

103	7	1.3	333	2	JC2329	translation initia	176	7	1.3	793	1	JH0628	caldesmon - human
104	7	1.3	336	2	T50318	conserved hypothe	177	7	1.3	794	2	T09643	glutamine-tRNA lig
105	7	1.3	339	2	F97190	phenylalanyl-tRNA	178	7	1.3	811	2	S76690	hypothetical prote
106	7	1.3	341	2	G86315	hypothetical prote	179	7	1.3	814	2	A70399	conserved hypothe
107	7	1.3	345	1	K2EC	alkaline phosphata	180	7	1.3	814	2	T02195	hypothetical prote
108	7	1.3	345	2	G91079	alkaline phosphata	181	7	1.3	824	2	G81748	conserved hypothe
109	7	1.3	345	2	H85924	alkaline phosphata	182	7	1.3	829	2	D71485	probable adenylate
110	7	1.3	346	2	AB3318	sodium-dependent p	183	7	1.3	846	1	VCLJND	env polypeptide pr
111	7	1.3	351	2	E71372	probable translati	184	7	1.3	865	2	A85032	hypothetical prote
112	7	1.3	356	2	JC4421	peptide chain rele	185	7	1.3	879	2	S49910	chloroplast outer
113	7	1.3	359	2	G84297	MAPK-activated pro	186	7	1.3	881	2	S67026	probable membrane
114	7	1.3	363	2	H81410	hydrogenase isoenz	187	7	1.3	883	2	T40128	kinesin-like prote
115	7	1.3	374	2	S46810	hypothetical prote	188	7	1.3	886	2	AB1785	the two components
116	7	1.3	376	2	S63613	probable ATP-bind	189	7	1.3	896	2	AF1409	the two components
117	7	1.3	378	2	T18486	hypothetical prote	190	7	1.3	907	2	I50404	p50B/p97 (Lyt-10)
118	7	1.3	381	2	C86358	Similar to zinc fi	191	7	1.3	924	2	T09220	exocyst complex pr
119	7	1.3	383	1	S15624	B2 protein - human	192	7	1.3	940	2	C84947	isoleucine-tRNA li
120	7	1.3	387	2	A12569	hypothetical prote	193	7	1.3	972	2	T39876	hypothetical prote
121	7	1.3	388	2	D72713	hypothetical prote	194	7	1.3	973	2	H86428	hypothetical prote
122	7	1.3	389	2	E81806	O-succinylhomoseri	195	7	1.3	990	2	T01896	probable Na+/K+/Cl
123	7	1.3	389	2	G81062	trans-sulfuration	196	7	1.3	1033	2	T42701	hypothetical prote
124	7	1.3	390	2	C81852	L-lactate dehydrog	197	7	1.3	1042	2	A57534	mucin 5AC (clone I
125	7	1.3	405	1	JH0795	calreticulin precu	198	7	1.3	1044	2	H71049	translation initia
126	7	1.3	420	2	AH2225	two-component sens	199	7	1.3	1057	2	H83273	ribonuclease B PA2
127	7	1.3	423	2	AD2455	hypothetical prote	200	7	1.3	1064	2	T44419	carbamoyl-phosphat
128	7	1.3	424	2	H84806	hypothetical prote	201	7	1.3	1065	1	I40169	carbamoyl-phosphat
129	7	1.3	424	2	S78610	hypothetical prote	202	7	1.3	1079	2	T38913	translation initia
130	7	1.3	425	2	T12473	hypothetical prote	203	7	1.3	1085	2	S62516	hypothetical coile
131	7	1.3	432	1	FGMGS	fibrinogen gamma c	204	7	1.3	1139	2	B54962	sterol regulatory
132	7	1.3	440	2	D95952	probable glycosyl	205	7	1.3	1141	2	A54962	sterol regulatory
133	7	1.3	446	2	B70543	hypothetical prote	206	7	1.3	1191	2	S27329	DNA topoisomerase
134	7	1.3	459	2	E75324	ARG8/DapE/AcyI fam	207	7	1.3	1192	1	ISXZAS	DNA topoisomerase
135	7	1.3	461	2	T14822	probable serine/th	208	7	1.3	1229	2	T25697	hypothetical prote
136	7	1.3	468	2	D95411	probable decarboxy	209	7	1.3	1271	2	A45355	glutamate rich pro
137	7	1.3	474	2	T39587	serine/threonine-p	210	7	1.3	1299	2	AH2090	two-component hybr
138	7	1.3	480	2	T08415	hypothetical prote	211	7	1.3	1353	2	T19157	probable metal bir
139	7	1.3	495	2	C95144	glucose-6-phosphat	212	7	1.3	1372	2	B71724	dna-directed RNA i
140	7	1.3	495	2	A98012	glucose-6-phosphat	213	7	1.3	1372	2	F97722	hypothetical prote
141	7	1.3	495	2	A44489	GT box-binding pro	214	7	1.3	1409	2	S74916	alkaline phosphata
142	7	1.3	496	2	C83122	probable aldehyde	215	7	1.3	1424	2	T19156	probable metal bir
143	7	1.3	506	2	C86164	P15K9.12 protein -	216	7	1.3	1503	2	T01098	chloroplast outer
144	7	1.3	507	2	G89908	Catalase (imported	217	7	1.3	1531	2	T48946	hypothetical prote
145	7	1.3	509	2	AH0578	citrate (pro-38)-l	218	7	1.3	1559	2	T07757	probable DNA (cyc
146	7	1.3	515	2	D71319	probable glucose-6	219	7	1.3	1657	2	A54854	Ras GTPase activat
147	7	1.3	553	2	G71543	probable fructose-	220	7	1.3	1927	2	A59236	embryonic muscle m
148	7	1.3	557	2	JC6539	tumor necrosis fac	221	7	1.3	1931	2	A59234	slow myosin heavy
149	7	1.3	562	2	S75308	DNA ligase (EC 6.5	222	7	1.3	1939	1	A46762	myosin heavy chain
150	7	1.3	564	2	E87631	acyl-CoA dehydroge	223	7	1.3	1940	1	S04090	myosin heavy chain
151	7	1.3	565	2	A87694	ribosomal protein	224	7	1.3	1940	1	A24922	hypothetical prote
152	7	1.3	570	2	D83964	fibronectin/fibrin	225	7	1.3	2019	2	T27702	hypothetical prote
153	7	1.3	571	2	S52276	translation initia	226	7	1.3	2123	2	P86348	hypothetical prote
154	7	1.3	572	2	G69977	fibronectin-bindin	227	7	1.3	2219	2	T27684	hypothetical prote
155	7	1.3	578	2	S28492	hypothetical prote	228	7	1.3	2541	2	T29340	hypothetical prote
156	7	1.3	578	2	E36892	Vi polysaccharide	229	7	1.3	3034	2	T14119	seven-pass transme
157	7	1.3	578	2	AC1041	hypothetical prote	230	7	1.3	3190	2	H83343	CRAB-binding prote
158	7	1.3	579	2	E71557	hypothetical prote	231	7	1.3	3432	2	T13828	probable non-ribos
159	7	1.3	579	2	JH0820	160K golgi antigen	232	7	1.3	4342	2	T08030	dynein beta heavy
160	7	1.3	584	2	E97146	indolepyruvate fer	233	7	1.3	4568	2	T04914	gamma2-gliadin p25
161	7	1.3	585	2	I58403	H4 protein - human	234	6	1.1	4589	2	S82125	G-alpha-1 protein
162	7	1.3	590	2	S57594	hypothetical prote	235	6	1.1	18	2	A32384	hypothetical prote
163	7	1.3	601	1	A64222	heat shock protein	236	6	1.1	31	2	F81079	hypothetical prote
164	7	1.3	620	2	G71413	hypothetical 7K pr	237	6	1.1	44	2	E70915	storage protein -
165	7	1.3	630	2	A41314	hypothetical prote	238	6	1.1	45	2	A24941	light-harvesting p
166	7	1.3	648	2	T23864	collagen alpha 3(I	239	6	1.1	49	1	L8RF8C	hypothetical prote
167	7	1.3	675	2	S20819	dnak-type molecula	240	6	1.1	51	2	AH1730	hypothetical prote
168	7	1.3	687	2	JN0667	probable membrane	241	6	1.1	51	2	T09177	ydaB protein - Bsc
169	7	1.3	703	2	AG0242	hypothetical prote	242	6	1.1	56	2	AD0739	probable bacterio
170	7	1.3	729	2	T50989	beta-N-acetylhexos	243	6	1.1	56	2	C90476	hypothetical prote
171	7	1.3	761	2	JC7821	myosin heavy chain	244	6	1.1	58	2	D97837	proline/betaine tr
172	7	1.3	784	2	I51302	carnitine O-palmit	245	6	1.1	63	2	D69115	hypothetical prote
173	7	1.3	772	2	G02860	carnitine palmitoy	246	6	1.1	63	2	AG3289	hypothetical prote
174	7	1.3	772	2	S65532	hypothetical prote	247	6	1.1	68	2	JU0099	hypothetical 7K pr
175	7	1.3	776	2	T19900		248	6	1.1				

ATTACHMENTS

JC7222
 77K muscle-derived protein - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
 C:Accession: JC7222
 C:Revised: 09-Jun-2000
 C:Author: Fujimori, K.; Kiyosue, K.; Nishimune, H.; Kasai, M.; Taguchi, T.; Ruyeda, A.; Fukui, I.; Blochem. Biophys. Res. Commun. 289, 564-569, 2000
 A:Title: MDP77: A novel neurite-outgrowth-promoting protein predominantly expressed in C
 A:Reference number: JC7222; PMID:20175243; PMID:10708594
 A:Accession: JC7222
 A:Molecule type: mRNA
 A:Residues: 1-676 <YE>
 A:Cross-references: GB:D89999; NID:G7619883; PIDN:BA94755.1; PID:G7619884
 A:Experimental source: crus muscle
 C:Comment: This protein, a glycoprotein and a neurite-outgrowth-promoting protein, is im

A:Reference number: Z20673
A:Accession: T29728
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-162 <DUZ>
A:Cross-references: EMBL:U55370; PIDN:AAA97999.1; GSPDB:GN00023; CESP:K03B4.7
A:Experimental source: strain Bristol N2; clone K03B4
C:Genetics:
A:Gene: CESP:K03B4.7
A:Map position: 5
A:introns: 58/3; 105/3

Query Match 1.5%; Score 8; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 290 AERLUKLI 297
|||
Db 89 AERLUKLI 96

RESULT 5
T32354
hypochemical protein C08E3.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32354
R:Miller, N.; Kramer, J.; Keppler, D.
A:Description: The sequence of C. elegans cosmid C08E3.
A:Reference number: Z21155
A:Accession: T32354
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-211 <ML>
A:Cross-references: EMBL:AF025457; PIDN:AA079965.1; GSPDB:GN00020; CESP:C08E3.4
A:Experimental source: strain Bristol N2; clone C08E3
C:Genetics:
A:Gene: CESP:C08E3.4
A:Map position: 2
A:introns: 36/3; 149/3

Query Match 1.5%; Score 8; DB 2; Length 211;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 290 AERLUKLI 297
|||
Db 81 AERLUKLI 88

RESULT 6
T52558
translation elongation factor eEF1alpha (clone 2) [validated] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 07-Jul-2003
C:Accession: T52558
R:Hericourt, F.; Jupin, I.
A:Title: Molecular cloning and characterization of the Arabidopsis thaliana alpha-subunit
A:Reference number: Z2614
A:Accession: T52558
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-224 <HER>
A:Cross-references: EMBL:AJ249597; PIDN:CAB64730.1
C:Genetics:
A:Gene: eEF1alpha2
A:Function:
A:Description: involved in translation elongation; able to complement a mutant yeast strain
C:Superfamily: translation elongation factor eEF-1 beta chain; translation elongation factor

Query Match 1.5%; Score 8; DB 2; Length 224;

Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 AAPAVEAZ 28
|||
Db 84 AAPAVEAZ 91

RESULT 7
E87236
probable membrane protein. [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: E87236
R:Coile, S.T.; Egilmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; F
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, E.; Holroyd
eam, M.A.; Rutherford, K.M.
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: E87236
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-229 <STO>
A:Cross-references: GB:AL450380; NID:G13093841; PIDN:CAC32147.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML2615
C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0200

Query Match 1.5%; Score 8; DB 2; Length 229;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 PSQAAPAV 25
|||
Db 200 PSQAAPAV 207

RESULT 8
S71315
deoxyguanosine kinase (EC 2.7.1.113) precursor - human
C:Species: Homo sapiens (man)
C:Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
C:Accession: S71315; S78432
R:Wang, L.; Hellman, U.; Eriksson, S.
FEBS Lett. 390, 39-43, 1996
A:Title: Cloning and expression of human mitochondrial deoxyguanosine kinase cDNA.
A:Reference number: S71315; MUID:96314545; PMID:8706825
A:Accession: S71315
A:Molecule type: mRNA
A:Residues: 1-260 <WAW>
A:Cross-references: EMBL:X97386; PIDN:CAA66054.1
A:Experimental source: tissue brain
R:Wang, L.; Hellman, U.; Eriksson, S.
submitted to the EMBL Data Library, April 1996
A:Description: Cloning and expression of human deoxyguanosine kinase cDNA.
A:Reference number: S78432
A:Accession: S78432
A:Molecule type: mRNA
A:Residues: 1-18, 'R', '20-260 <WAW>
A:Cross-references: EMBL:X97386; PIDN:CAA66054.1
A:Experimental source: brain
C:Genetics:
A:Gene: dgk
A:Genome: nuclear
C:Superfamily: human deoxycytidine kinase
C:Keywords: mitochondrion; phosphotransferase
P:1-22/Domain: transit peptide (mitochondrion) #status predicted <TMP>
E:23-260/Product: transit peptide (mitochondrion)
P:125-132/Region: DRS motif
F:185-191/Region: arginine-rich

Query Match 1.5%; Score 8; DB 2; Length 260;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Gaps 0;

QY 245 QRAPEE 252
 |||||
 Db 188 QRAPEE 195

RESULT 9
 C88638
 Protein F58F6.1 [imported] - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 15-Sep-2003
 C;Accession: C88638
 R;Anonymous, The C. elegans Sequencing Consortium.
 Science 283, 2012-2018, 1998
 A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A;Reference number: A75000; MUID:99069613; PMID:9851916
 A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
 A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A;Accession: C88638
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-281 <STO>
 A;Cross-references: GB:chr_IV; PIDN:AAB88358.1; PID:g2662599; GSPDB:GN00022; CESP:F58F6.
 C;Genetics:
 A;Gene: F58F6.1
 A;Map position: 4

Query Match 1.5%; Score 8; DB 2; Length 281;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 GGPEDGA 84
 |||||
 Db 241 GGPEDGA 248

RESULT 10
 F03985
 Hypothetical protein F6G3.120 - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
 C;Accession: T08985
 R;Savan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, May 1999
 A;Reference number: Z16520
 A;Accession: F03985
 A;Molecule type: DNA
 A;Residues: 1-312 <BEV>
 A;Cross-references: EMBL:AL078464; GSPDB:GN00062; ATSP:F6G3.120
 A;Experimental source: cultivar Columbia; BAC clone F6G3
 C;Genetics:
 A;Gene: ATSP:F6G3.120
 A;Map position: 4
 A;Introns: 55/1; 94/3; 115/3; 151/3; 168/3; 189/3; 241/3; 264/3; 285/2

Query Match 1.5%; Score 8; DB 2; Length 312;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 QQAQEMLK 333
 |||||
 Db 240 QQAQEMLK 247

RESULT 11
 E71909
 Hypothetical protein jhp0612 - Helicobacter pylori (strain J99)
 C;Species: Helicobacter pylori
 A;Variety: strain J99
 C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C;Accession: E71909
 R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Werberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.P.;
 Nature 397, 176-180, 1999
 A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric pat
 A;Reference number: A71800; MUID:99120557; PMID:9923682
 A;Accession: E71909
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-450 <ARN>
 A;Cross-references: GB:AE001493; GB:AE001439; NID:g4155161; PIDN:AAD06193.1; PID:g41551
 A;Experimental source: strain J99
 C;Genetics:
 A;Gene: jhp0612

Query Match 1.5%; Score 8; DB 2; Length 450;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 LEDILSTY 70
 |||||
 Db 283 LEDILSTY 290

RESULT 12
 D88710
 Protein C43G2.1 [imported] - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
 C;Accession: D88710
 R;Anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo
 A;Reference number: A75000; MUID:99069613; PMID:9851916
 A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_el
 A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103 1999; an
 A;Accession: D88710
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-467 <STO>
 A;Cross-references: GB:chr_IV; PIDN:AAB09107.1; PID:g1572756; GSPDB:GN00022; CESP:C43G2
 C;Genetics:
 A;Gene: C43G2.1
 A;Map position: 4

Query Match 1.5%; Score 8; DB 2; Length 467;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 LAALCKXY 180
 |||||
 Db 22 LAALCKXY 29

RESULT 13
 H97769
 Hypothetical protein murD [imported] - Rickettsia conorii (strain Malish 7)
 C;Species: Rickettsia conorii
 C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
 C;Accession: H97769
 R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Pournier, P.E.; Barbe, V.; Samson, D.; R
 Science 293, 2093-2098, 2001
 A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
 A;Reference number: A97700; MUID:2142074; PMID:11557993
 A;Accession: H97769
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-500 <KUR>
 A;Cross-references: GB:AE006914; PIDN:AAL03098.1; PID:g15619640; GSPDB:GN00173
 C;Genetics:
 A;Gene: murD
 C;Superfamily: UDP-N-acetylmuramate-alanine ligase

Query Match 1.5%; Score 8; DB 2; Length 500;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 85 QGPEAPE 92
Db 278 QGPEAPE 285

RESULT 14
H75514
glutamyl-tRNA synthetase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: H75514
R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Shen, M.; Vamathevan, J.G.; Lam, P.; McDonald, L.; Otterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: H75514
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-525 <WHI>
A:Cross-references: GB:AB001907; GB:AS000513; NID:96458162; PIDN:AAFI0063.1; PID:9645817
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0485
A:Map position: 1
C:Superfamily: glutamate-tRNA ligase; glutamine-tRNA ligase homology

Query Match 1.5%; Score 8; DB 2; Length 525;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 496 APSSPRT 503
Db 45 APSSPRT 52

RESULT 15
A55887
caldesmon, non-muscle - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 13-Aug-1999
C:Accession: A55887
R:Yamashiro, S.; Yamakita, Y.; Yoshida, K.; Takiguchi, K.; Matsumura, F.
J. Biol. Chem. 270, 4023-4030, 1995
A:Title: Characterization of the COOH terminus of non-muscle caldesmon mutants lacking
A:Reference number: A55887; MUID:95181370; PMID:7876150
A:Accession: A55887
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-531 <YAM>
A:Cross-references: GB:U18419; NID:G622966; PIDN:AAA68521.1; PID:G622967
A:Note: authors translated the codon GCC for residue 68 as Val
C:Superfamily: caldesmon
C:Keywords: phosphoprotein

Query Match 1.5%; Score 8; DB 2; Length 531;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 247 AREEEKR 254
Db 334 AREEEKR 341

RESULT 16
C36811
hypothetical protein ORF50 - saimirine herpesvirus 1 (strain 11)
C:Species: saimirine herpesvirus 1

A:Note: host Saimiri sciureus (common squirrel monkey)
C:Date: 16-Oct-1992 #sequence_revision 16-Oct-1992 #text_change 27-Jan-1995
C:Accession: C36811
R:Albrecht, J.
submitted to the EMBL Data Library, January 1992
A:Description: Primary structure of the herpesvirus saimir genome.
A:Reference number: A36806
A:Accession: C36811
A:Molecule type: DNA
A:Residues: 1-535 <ALB>
A:Cross-references: GB:X64346
R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.; W
J. Virol. 66, 5047-5058, 1992
A:Title: Primary structure of the herpesvirus saimir genome.
A:Reference number: A37309; MUID:92333688; PMID:1321287
A:Contents: annotation; protein-coding frames
A:Note: neither protein nor nucleotide sequence is given
C:Genetics:
A:Gene: 50

Query Match 1.5%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 GLGKBITL 158
Db 187 GLGKBITL 194

RESULT 17
A37994
RFl protein - saimirine herpesvirus 1
C:Species: saimirine herpesvirus 1
C:Date: 23-Aug-1991 #sequence_revision 30-Jan-1993 #text_change 04-Mar-1994
C:Accession: A37994
R:Nicholas, J.; Coles, L.S.; Newman, C.; Honess, R.W.
J. Virol. 65, 2457-2466, 1991
A:Title: Regulation of the herpesvirus saimir (HVS) delayed-early 110-kilodalton prom
A:Reference number: A37994; MUID:91202588; PMID:1850023
A:Accession: A37994
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-535 <NIC>
A:Cross-references: GB:M60850

Query Match 1.5%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 GLGKBITL 158
Db 187 GLGKBITL 194

RESULT 18
E90364
hypothetical protein SSO1981 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: E90364
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,
arrett, R.A.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: E90364
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-568 <GR>
A:Cross-references: GB:AE006641; NID:gl3815258; PIDN:AAK42172.1; GSPDB:GN0015;
C:Genetics:
A:Gene: SSO1981

```

Query Match      1.5%; Score 8; DB 2; Length 568;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLEK 411
Db 91 KKIKKLEK 98
|||||

RESULT 19
S62141
transcription initiation factor IIIB chain B'' - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein N2682; protein YNL039W
C:Species: Saccharomyces cerevisiae
C>Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 21-Jul-2000
C:Accession: S62141; S62961; S62912
R:Kasavatis, G.A.; Nguyen, S.T.; Kobayashi, R.; Kumar, A.; Geiduschek, E.P.; Pisano, M.
Proc. Natl. Acad. Sci. U.S.A. 92, 9786-9790, 1995
A:Title: Cloning, expression, and function of TFC5, the gene encoding the B'' component
A:Reference number: S62141; PMID:7568218
A:Accession: S62141
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-594 <KAS>
A:Cross-references: EMBL:U31819; NID:G1019114; PIDN:AAC49073.1; PID:G1019115
R:Duesterhoft, A.; Floeth, M.; Fritz, C.; Heuss-Neitzel, D.; Hilbert, H.; Moestl, D.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S62944
A:Accession: S62961
A:Molecule type: DNA
A:Residues: 1-594 <DUE>
A:Cross-references: EMBL:W71315; NID:G1301889; PID:e239882; PID:G1301890; MIPS:YNL039W
A:Experimental source: strain S288C
R:Rueh, J.; Conesa, C.; Dieci, G.; Lefebvre, O.; Duesterhoft, A.; Ottoneillo, S.; Sente
EMBO J. 15, 1941-1949, 1996
A:Title: A suppressor of mutations in the class III transcription system encodes a compo
A:Reference number: S62912; PMID:96203116; PMID:8617241
A:Accession: S62912
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-594 <RUE>
A:Cross-references: EMBL:U38415; NID:G1145357; PIDN:AAC49364.1; PID:G1145358
C:Genetics:
A:Gene: SGD:TFC5
A:Cross-references: SGD:S0004984; MIPS:YNL039W
A:Map position: 14L
C:Keywords: transcription initiation

Query Match      1.5%; Score 8; DB 2; Length 594;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEERKE 256
Db 334 EEEERKE 341
|||||

RESULT 20
S65169
hypothetical protein YPL158c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein P2570
C:Species: Saccharomyces cerevisiae
C>Date: 10-Dec-1994 #sequence_revision 31-May-1996 #text_change 19-Apr-2002
C:Accession: S65169; S69441
R:Purnelle, B.; Coster, F.; Goffeau, A.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S65154
A:Accession: S65169
A:Molecule type: DNA
A:Residues: 1-758 <PUR>
A:Cross-references: EMBL:Z73514; NID:G1370335; PID:e247047; PID:G1370336; MIPS:YPL158c
A:Experimental source: strain S288C (AB972)

R:Purnelle, B.; Comblez, S.; Coster, F.; Naveau, F.; Goffeau, A.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of 55 kb on the left arm of yeast chromosome XVI identifies
ogues to the human phosphotyrosyl phosphatase activator PTPA and a homologue to the plan
A:Reference number: S69428
A:Accession: S69441
A:Molecule type: DNA
A:Residues: 1-758 <PUW>
A:Cross-references: EMBL:X96770; NID:G1403537; PID:e239043; PID:G1403551
C:Genetics:
A:Cross-references: SGD:S0006079
A:Map position: 16L

Query Match      1.5%; Score 8; DB 2; Length 758;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEERKE 256
Db 687 EEEERKE 694
|||||

RESULT 21
T32183
hypothetical protein K09C6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32183
R:Goela, D.; Harper, M.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid K09C6.
A:Reference number: Z21131
A:Accession: T32183
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-761 <GOE>
A:Cross-references: EMBL:AF022975; PIDN:AAB70668.1; GSPDB:GN00023; CESP:K09C6.1
A:Experimental source: strain Bristol N2; clone K09C6
C:Genetics:
A:Gene: CESP:K09C6.1
A:Map position: 5
A:Introns: 48/3; 91/3; 123/2; 146/2; 200/2; 238/3; 269/1; 302/1; 481/3; 591/1; 639/1

Query Match      1.5%; Score 8; DB 2; Length 761;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 AELLEER 188
Db 710 AELLEER 717
|||||

RESULT 22
T52292
endopeptidase Clp (EC 3.4.21.92) ATP-binding chain C, chloroplast [imported] - Arabidop
N:Alternate names: ATP-dependent clp proteinase chain C
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 01-Dec-2000
C:Accession: T52292
R:Clarke, A.K.
submitted to the EMBL Data Library, September 1997
A:Description: A cDNA clone coding for the chloroplast ClpC protein from Arabidopsis th
A:Reference number: Z26020
A:Accession: T52292
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-928 <CUA>
A:Cross-references: EMBL:AF022909; PIDN:AAC04687.1
C:Superfamily: endopeptidase Clp ATP-binding chain
C:Keywords: ATP; chloroplast; hydrolase; nucleotide binding; serine proteinase

Query Match      1.5%; Score 8; DB 2; Length 928;
Best Local Similarity 100.0%; Pred. No. 35;

```

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 EEIRQSDE 134
|||||
Db 359 EEIRQSDE 366

RESULT 23

T52456
endopeptidase Clp ATP-binding chain C [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 08-Dec-2000
C:Accession: T52456
R:Nakabayashi, K.; Ito, M.; Kiyosue, T.; Shinozaki, K.; Watanabe, A.
Plant Cell Physiol. 40, 504-514, 1999
A:Title: Identification of clp genes expressed in senescing Arabidopsis leaves.
A:Reference number: Z2128; MUID:99356780; PMID:10427773
A:Accession: T52456
A:Status: preliminary;
A:Molecule type: mRNA
A:Residues: 1-952 <NAK>
A:Cross-references: EMBL:AB022324; NID:G5360573; PIDN:BA82062.1; PID:G5360574
A:Experimental source: strain Columbia
C:Genetics:
A:Gene: clpC
C:Superfamily: endopeptidase Clp ATP-binding chain

Query Match 1.5%; Score 8; DB 2; Length 952;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 EEIRQSDE 134
|||||
Db 381 EEIRQSDE 388

RESULT 24

T49283
AtClpC - Arabidopsis thaliana
N:Alternate names: protein T21J18.140
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Feb-2001
C:Accession: T49283
R:Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Rudd, S.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25021
A:Accession: T49283
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-952 <R1E>
A:Cross-references: EMBL:AL132963; GSPDB:GN00061; ATSP:T21J18.140
A:Experimental source: cultivar Columbia; BAC clone T21J18
C:Genetics:
A:Gene: ATSP:T21J18.140
A:Map position: 3
A:Introns: 212/1; 250/3; 353/3; 425/3; 498/2; 523/3; 553/3
C:Superfamily: endopeptidase Clp ATP-binding chain
C:Keywords: ATP; molecular chaperone; nucleotide binding; P-loop
F:323-330/Region: nucleotide-binding motif A (P-loop)
F:390-395/Region: nucleotide-binding motif B
F:666-673/Region: nucleotide-binding motif A (P-loop)
F:734-739/Region: nucleotide-binding motif B
F:329/Binding site: ATP (Lys) #status predicted
F:672/Binding site: ATP (Lys) #status predicted

Query Match 1.5%; Score 8; DB 2; Length 952;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 EEIRQSDE 134
|||||
Db 381 EEIRQSDE 388

RESULT 25

AE3243
conjugal transfer protein, Dtr system traA [imported] - Agrobacterium tumefaciens (stra
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AE3243
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
star, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AE3243
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1100 <KUR>
A:Cross-references: GB:AE008690; PIDN:AAL46363.1; PID:gl7744153; GSPDB:GN00169
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: traA
A:Genome: plasmid

Query Match 1.5%; Score 8; DB 2; Length 1100;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 135 VGDRDHR 142
|||||
Db 671 VGDRDHR 678

RESULT 26

T03419
traA protein - Agrobacterium tumefaciens plasmid pTiC58
C:Species: Agrobacterium tumefaciens
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C:Accession: T03419
R:Piper, K.R.; Beck von Bodman, S.; Cook, D.M.; Hwang, I.; Kim, H.; Farrand, S.K.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z14943
A:Accession: T03419
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1101 <PIP>
A:Cross-references: EMBL:AF010180; NID:g3153171; PIDN:AAC17212.1; PID:g1103910
C:Genetics:
A:Gene: traA
A:Genome: plasmid pTiC58

Query Match 1.5%; Score 8; DB 2; Length 1101;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 135 VGDRDHR 142
|||||
Db 672 VGDRDHR 679

RESULT 27

C97702
cell surface antigen [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: C97702
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R.
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: C97702
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-1902 <KUR>
C:Date: 06-Oct-1994 #sequence_revision 26-Jul-1996 #text_change 15-Oct-1999
C:Genetics: 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
A:Gene: scal

Query Match
Best Local Similarity 100.0%; Pred. No. 68;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 KIKLEKE 412
|||||
DB 356 KIKLEKE 363
|||||

RESULT 28
T18519
myosin X - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Feb-2001
C:Accession: T18519
R:Corey, D.P.; Derfler, B.H.; Solc, C.K.; Duyk, G.M.; Cheney, R.E.
submitted to the EMBL Data Library, April 1996
A:Description: Cloning and expression of myosin X, a novel unconventional myosin with pI
A:Reference number: Z18942
A:Accession: T18519
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2052 <COR>
A:Cross-references: EMBL:U55042; NID:G1755048; PID:G1755049; PIDN:AA839486.1
A:Experimental source: aorta
C:Superfamily: myosin motor domain homology
C:Keywords: nucleotide binding; P-loop
P:66-727/Domain: myosin motor domain homology <MWO>
F:157-164/Region: nucleotide-binding motif A (P-loop)

Query Match
Best Local Similarity 100.0%; Pred. No. 73;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 REEEKKK 255
|||||
DB 824 REEEKKK 831
|||||

RESULT 29
S45032
homeoic protein SPOX TAL - Tethya aurantia (fragment)
C:Species: Tethya aurantia
C:Date: 06-Oct-1994 #sequence_revision 26-Jul-1996 #text_change 15-Oct-1999
C:Accession: S45032
R:Degnan, B.M.; Degnan, S.M.; Giusti, A.; Morse, D.E.
submitted to the EMBL Data Library, May 1994
A:Description: A Hox homeobox gene in sponges.
A:Reference number: S45030
A:Accession: S45032
A:Molecule type: DNA
A:Residues: 1-23 <DEG>
A:Cross-references: EMBL:X79265; NID:G488780; PIDN:CAA5852.1; PID:G495148
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 ELAERLK 294
|||||
DB 12 ELAERLK 18
|||||

RESULT 30
B91253

hypothetical protein EC84994 [imported] - Escherichia coli (strain O157:H7, substrain R
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: B91253
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B91253
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-60 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA838417.1; PID:G13364470; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 050952
C:Genetics: 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
A:Gene: EC84994

Query Match
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AQAPSSP 500
|||||
DB 48 AQAPSSP 54
|||||

RESULT 31
AG1905
hypothetical protein asl0793 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG1905
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, I.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1905
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-62 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA872750.1; PID:G17130138; GSPDB:GN00175
A:Experimental source: strain PCC 7120
C:Genetics: 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
A:Gene: asl0793

Query Match
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 DLSAGGQ 477
|||||
DB 21 DLSAGGQ 27
|||||

RESULT 32
AC1755
hypothetical protein lin2594 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AC1755
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitourram, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1755
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-75 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC97811.1; PID:g16415106; GSPDB:GN00178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin2584

Query Match 1.3%; Score 7; DB 2; Length 75;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 LKXLEQ 299
DB 58 LKXLEQ 64
|||||

RESULT 33
D75326
hypothetical protein - Deinococcus radiodurans (strain RL)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
A:Accession: D75326
R:White, O.; Eisen, J.A.; Heidelberg, J.P.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans RL.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: D75326
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-79 <WHI>
A:Cross-references: GB:AE002038; GB:AE000513; MID:G6459790; PIDN:AAF11565.1; PID:G645980
A:Experimental source: strain RL
C:Genetics:
A:Gene: DR2007
A:Map position: 1

Query Match 1.3%; Score 7; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ELGRQLE 64
DB 12 ELGRQLE 18
|||||

RESULT 34
S78290
ribosomal protein L21, chloroplast - Odontella sinensis chloroplast
C:Species: chloroplast Odontella sinensis
C:Date: 17-Feb-1998 #sequence_revision 26-Feb-1998 #text_change 20-Jun-2000
A:Accession: S78290
R:Kowalik, K.V.; Stoebe, B.; Schaffran, I.; Kroth-Pancic, P.; Freier, U.
Plant Mol. Biol. Rep. 13, 336-342, 1995
A:Title: The Chloroplast Genome of a chlorophyll a+c- containing Alga, Odontella sinensis
A:Reference number: S78238
A:Accession: S78290
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-105 <ROW>
A:Cross-references: EMBL:267753; MID:g1185127; PIDN:CAA91663.1; PID:g1185180
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995
C:Genetics:
A:Gene: rpl21
A:Genome: chloroplast
C:Superfamily: Escherichia coli ribosomal protein L21
C:Keywords: chloroplast; protein biosynthesis; ribosome

Query Match 1.3%; Score 7; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LGKBITL 158
DB 30 LGKBITL 36
|||||

RESULT 35
S10587
cystatin C - rat
C:Species: Rattus sp. (rat)
C:Date: 21-Nov-1993 #sequence_revision 03-Nov-1995 #text_change 16-Jul-1999
A:Accession: S10587
R:Esmond, P.; Ennard, A.; Faucher, D.; Capony, J.P.; Derancourt, J.; Brillard, M.; Gaut
Biol. Chem. Hoppe-Seyler 371 (Suppl.), 161-166, 1990
A:Title: Rat cystatin C: the complete amino acid sequence reveals a site for N-glycosyl
A:Reference number: S10587; MUID:90380276; PMID:2400577
A:Accession: S10587
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-120 <ESN>
A:Note: 43-Asn was also found
A:Note: the sequence from Fig. 2 is inconsistent with that from Fig. 1 in having 18-Ala
C:Superfamily: cystatin; cystatin homology
P:9-120/Domain: cystatin homology <CYS>

Query Match 1.3%; Score 7; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 BEGVORA 247
DB 20 BEGVORA 26
|||||

RESULT 36
T44473
conserved hypothetical protein tnpG [imported] - Shigella flexneri
C:Species: Shigella flexneri
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
A:Accession: T44473
R:Yoss, J.E.; Cardozo, T.J.; Zychlinsky, A.; Groisman, E.A.
Mol. Microbiol. 33, 74-83, 1999
A:Title: The selC-associated SHI-2 pathogenicity island of Shigella flexneri.
A:Reference number: Z22779; MUID:99340540; PMID:10411725
A:Accession: T44473
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-121 <WOS>
A:Cross-references: EMBL:AF141323; MID:G5532445; PIDN:RAD4740.1; PID:G5532456
A:Experimental source: strain M90T; serotype 5a
C:Genetics:
A:Gene: tnpG
C:Superfamily: Agrobacterium tumefaciens insertion sequence IS1312 hypothetical protein

Query Match 1.3%; Score 7; DB 2; Length 121;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 LLKEAVE 352
DB 96 LLKEAVE 102
|||||

RESULT 37
S07085
cystatin C precursor - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 01-Dec-1993 #sequence_revision 03-Aug-1995 #text_change 16-Jul-1999
A:Accession: S07085; S01337; S21109
R:Cole, T.; Dickson, P.W.; Ennard, P.; Averill, S.; Risbridger, G.P.; Gauthier, F.; Sch
Eur. J. Biochem. 186, 35-42, 1989
A:Title: The cDNA structure and expression analysis of the genes for the cysteine prote
A:Reference number: S07085; MUID:90092122; PMID:2689174
A:Accession: S07085

A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-127 <COL>
A:Cross-references: EMBL:X16957; NID:G56041; PIDN:CAA34831.1; PID:G736290
R:Ennard, A.; Ennard, F.; Paucher, D.; Gauthier, F.
FEBS Lett. 236, 475-478, 1988
A>Title: Two rat homologues of human cystatin C.
A:Reference number: S01337; MUID:88313020; PMID:3044831
A:Molecule type: protein
A:Residues: 8-49 <PSN>
R:Ennard, A.; Ennard, F.; Guillou, F.; Gauthier, F.
FEBS Lett. 300, 131-135, 1992
A>Title: Production of the cysteine proteinase inhibitor cystatin C by rat Sertoli cells
A:Reference number: S21109; MUID:92255121; PMID:1563513
A:Accession: S21109
A:Molecule type: protein
A:Residues: 8, 'XX', 11-20 <ES2>
C:Superfamily: cystatin; cystatin homology
C:Keywords: cysteine proteinase inhibitor
F:16-127/Domain: cystatin homology <CR>
F:80-90,104-124/Disulfide bonds: #status predicted

Query Match 1.3%; Score 7; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 EGVQRA 247
|||||
DB 27 EGVQRA 33

RESULT 38
Hypothetical protein [imported] - Thermomonospora fusca
C:Species: Thermomonospora fusca
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
R:Irwin, D.I.; Zhang, S.; Wilson, D.B.
Submitted to the EMBL Data Library, April 1999
A:Description: Characterization of a Thermomonospora fusca family 48 exocellulase B6.
A:Reference number: Z22783
A:Accession: T44497
A>Status: preliminary; translated from GB/EMBL/DDB
A:Molecule type: DNA
A:Residues: 1-128 <IRW>
A:Cross-references: EMBL:AF144563; PIDN:AAD39948.1
A:Experimental source: strain YX

Query Match 1.3%; Score 7; DB 2; Length 128;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SSPGQPE 8
|||||
DB 108 SSPGQPE 114

RESULT 39
Major allergen Par j I (clone Pb) - Parietaria judaica
C:Species: Parietaria judaica
C>Date: 21-Apr-1997 #sequence_revision 18-Jul-1997 #text_change 02-Sep-2000
C:Accession: S77948; S43682; S41944
R:Geraci, D.
Submitted to the EMBL Data Library, September 1995
A:Description: cDNA cloning, expression and primary structure of Par j I, a major allergen
A:Reference number: S77948
A:Accession: S77948
A:Molecule type: mRNA
A:Residues: 1-133 <GER>
A:Cross-references: EMBL:X77414; NID:G992611; PIDN:CAA54587.1; PID:G992612
A:Experimental source: clone Pb

A>Note: this is a revision to the sequence from reference S43682
R:Costa, M.A.; Colombo, P.; Izzo, V.; Kennedy, K.; Venturilla, S.; Cocchiara, R.; Mistr
FEBS Lett. 341, 182-186, 1994
A>Title: cDNA cloning, expression and primary structure of Par j I, a major allergen of
A:Reference number: S43682; MUID:94185791; PMID:8137937
A:Accession: S43682
A:Molecule type: mRNA
A:Residues: 1-110, 'RSRPPTKGWRDP', 124, 'LSFRP', 130, 'HRKKPNDAFSTLG' <COS>
A:Cross-references: EMBL:X77414
A>Note: the authors translated the codon GTT for residue 88 as Leu
A>Note: this sequence has been revised in reference S77948
C:Genetics:
A:Gene: Pmai
C:Superfamily: phospholipid transfer protein

Query Match 1.3%; Score 7; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 EKEPSKG 122
|||||
DB 16 EKEPSKG 22

RESULT 40
B65240
Hypothetical protein, 13.4K - Escherichia coli (strain K-12) insertion sequence IS2
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 01-Mar-2002
C:Accession: B65240; F64891; H64763; B65092; S56497; JQ0039
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: B65240
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BLAT>
A:Cross-references: GB:AB000498; GB:U00096; NID:G2367368; PIDN:AAC77228.1; PID:G1790723
A:Experimental source: strain K-12, substrain MG1655
A:Genetics: Y21
A:Accession: F64891
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BLA2>
A:Cross-references: GB:AE000237; GB:U00096; NID:G1787665; PIDN:AAC74485.1; PID:G1787670
A:Experimental source: strain K-12, substrain MG1655
A:Genetics: Y22
A:Accession: H64763
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BLA3>
A:Cross-references: GB:AE000143; GB:U00096; NID:G1786554; PIDN:AAC73463.1; PID:G1786557
A:Experimental source: strain K-12, substrain MG1655
A:Genetics: Y11
A:Accession: B65092
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 16-136 <BLA4>
A:Cross-references: GB:AB000386; GB:U00096; NID:G2367187; PIDN:AAC76080.1; PID:G1789423
A:Experimental source: strain K-12, substrain MG1655
A:Genetics: Y25
A:Accession: F65069
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 16-136 <BLA5>
A:Cross-references: GB:AB000369; GB:U00096; NID:G2367168; PIDN:AAC75900.1; PID:G1789225,
A:Experimental source: strain K-12, substrain MG1655
A:Genetics: Y24
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A>Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92

A:Reference number: S56314; MUID:95334362; PMID:7610040
A:Accession: S56497
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BUR>
A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97168.1; PID:g537113
A:Genetics: Y26
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R:Ronecker, H.J.; Rak, B.
Gene 59, 291-296, 1987
A:Title: Genetic organization of insertion element IS2 based on a revised nucleotide sequence
A:Reference number: A91582; MUID:98137965; PMID:2630172
A:Accession: J00039
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 16-136 <RON>
A:Cross-references: GB:V00279; GB:J01732; GB:M18426; NID:g41525; PIDN:CAA23542.1; PID:g41525
A:Experimental source: K12, strain HB101
A:Genetics: HB1
C:Genetics: <Y21>
A:Gene: Y121
A:Mobile element: insertion sequence IS2
C:Genetics: <Y22>
A:Gene: Y121_2
A:Mobile element: insertion sequence IS2
C:Genetics: <Y11>
A:Gene: Y121_1
A:Mobile element: insertion sequence IS2
C:Genetics: <Y25>
A:Gene: Y121_5
A:Mobile element: insertion sequence IS2
C:Genetics: <Y24>
A:Gene: Y121_4
A:Mobile element: insertion sequence IS2
C:Genetics: <Y26>
A:Mobile element: insertion sequence IS2
C:Genetics: <HB1>
A:Mobile element: insertion sequence IS2
A:Start codon: GAG
C:Superfamily: Agrobacterium tumefaciens insertion sequence IS1312 hypothetical protein

Query Match 1.3%; Score 7; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 LLKEAVE 352
DB 111 LLKEAVE 117
|||||

RESULT 41
D64964
Hypothetical protein - Escherichia coli (strain K-12) insertion sequence IS2
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: D64964
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: D64964
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BLAT>
A:Cross-references: GB:AE000291; GB:U00096; NID:g1788298; PIDN:AACT5058.1; PID:g1788306;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: Y121_3
A:Mobile element: insertion sequence IS2
C:Superfamily: Agrobacterium tumefaciens insertion sequence IS1312 hypothetical protein

Query Match 1.3%; Score 7; DB 2; Length 136;

Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 LLKEAVE 352
DB 111 LLKEAVE 117
|||||

RESULT 42
S52933
major allergen Par j (clone PA2) - Parietaria judaica
C:Species: Parietaria judaica
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 02-Sep-2000
C:Accession: S52933
R:Geraci, D.
submitted to the EMBL Data Library, February 1995
A:Reference number: S52933
A:Accession: S52933
A:Molecule type: mRNA
A:Residues: 1-138 <GER>
A:Cross-references: EMBL:X85012; NID:g706810; PIDN:CAA59370.1; PID:g706811
A:Experimental source: clone PA2
C:Superfamily: phospholipid transfer protein

Query Match 1.3%; Score 7; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 EKEPSKG 122
DB 58 EKEPSKG 64
|||||

RESULT 43
A90258
conserved hypothetical protein [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: A90258
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: A90258
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-138 <KUR>
A:Cross-references: GB:AE006641; NID:g13814244; PIDN:AAK41320.1; GSPDB:GN00155
C:Genetics:
A:Gene: SSO1056

Query Match 1.3%; Score 7; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 LKAEER 338
DB 111 LKAEER 117
|||||

RESULT 44
G98112
hypothetical protein spr1930 [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: G98112
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: G98112
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAL00732.1; PID:G15459627; GSPDB:GN00174
C:Genetics:
A:Gene: spr1930

Query Match 1.3%; Score 7; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 404 KKIKKLE 410
Db 69 KKIKKLE 75

RESULT 45
B95248
Hypothetical protein SP2120 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: B95248
R:Tetreltin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
Nelson, I.; Hackey, E.K.; Holt, I.E.
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B95248
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-155 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK76179.1; PID:G14973632; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP2120

Query Match 1.3%; Score 7; DB 2; Length 155;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 404 KKIKKLE 410
Db 70 KKIKKLE 76

RESULT 46
T49567
related to attachment protein [imported] - Neurospora crassa
N:Alternate names: protein B208.150
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49567
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25022
A:Accession: T49567
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-158 <SCH>
A:Cross-references: EMBL:AL355930; GSPDB:GN00116; NCSP:B208.150
A:Experimental source: BAC clone B208; strain OR74A
C:Genetics:
A:Gene: NCSP:B208.150
A:Map position: 6

Query Match 1.3%; Score 7; DB 2; Length 158;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 240 KEEGVOR 246
Db 62 KEEGVOR 68

RESULT 47
T32043
Hypothetical protein K07E8.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32043
R:Jones, K.; Kramer, J.
A:Submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid K07E8.
A:Reference number: Z21116
A:Accession: T32043
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-159 <JON>
A:Cross-references: EMBL:AF016678; PIDN:AAB66151.1; GSPDB:GN00020; CESP:K07E8.6
A:Experimental source: strain Bristol N2; clone K07E8
C:Genetics:
A:Gene: CESP:K07E8.6
A:Map position: 2

Query Match 1.3%; Score 7; DB 2; Length 159;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 170 EEKLAAL 176
Db 54 EEKLAAL 60

RESULT 48
E71183
VPS29-like phosphoesterase-related protein PH1746 [similarity] - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 24-Aug-2001
C:Accession: E71183
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguch
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:96344137; PMID:9679194
A:Accession: E71183
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-163 <KAW>
A:Cross-references: GB:AP000007; MUID:G3236134; PIDN:BAA30860.1; PID:G3258177
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1746
C:Superfamily: human vacuolar protein-sorting protein VPS29 homolog; phosphoesterase co

Query Match 1.3%; Score 7; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 153 GKBITLL 159
Db 123 GKBITLL 129

RESULT 49
A69454
Hypothetical protein AF1634 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: A69454
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

```

.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.P.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: A69454
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-172 <KLE>
A:Cross-references: GB:AE000989; GB:AE000782; MID:g2689312; PIDN:AA89617.1; PID:g264892

Query Match      1.3%; Score 7; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 GKXITLL 159
DB 78 GKXITLL 84

RESULT 50
AE2913
Conserved hypothetical protein Atu2742 [imported] - Agrobacterium tumefaciens (strain C58)
A:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AE2913
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, B.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AE2913
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-176 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL43723.1; PID:g17741254; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
A:Genetics:
A:Gene: Atu2742
A:Map position: circular chromosome

Query Match      1.3%; Score 7; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 VSEELSR 61
DB 43 VSEELSR 49

RESULT 51
T37212
hypothetical protein Y57G7A.2 - Caenorhabditis elegans
A:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Aug-2000
C:Accession: T37212
R:Ozersky, P.
submitted to the EMBL Data Library, March 1999
A:Description: The sequence of C. elegans cosmid Y57G7A.
A:Reference number: 221634
A:Accession: T37212
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-177 <OZE>
A:Cross-references: EMBL:AF077542; PIDN:AAC26297.1
A:Genetics:
A:Map position: 2
A:Introns: 105/1; 133/1

```

```

A:Note: Y57G7A.2
C:Superfamily: Caenorhabditis elegans hypothetical protein Y57G7A.2

```

```

Query Match      1.3%; Score 7; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 424 NKALLEM 430
DB 109 NKALLEM 115

```

RESULT 52

```

C85165
Hypothetical protein dl3556w [imported] - Arabidopsis thaliana
A:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: C85165
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spri
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: AB5001; MUID:20083488; PMID:10617198
A:Accession: C85165
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-179 <STO>
A:Cross-references: GB:NC_001268; MID:g5302779; PIDN:CAB46057.1; GSPDB:GN00140
C:Genetics:
A:Gene: dl3556w
A:Map position: 4

```

```

Query Match      1.3%; Score 7; DB 2; Length 179;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 250 BEEKKEK 256
DB 108 BEEKKEK 114

```

RESULT 53

```

AB3445
H+-transporting two-sector ATPase (EC 3.6.3.14) [imported] - Brucella melitensis (strai
A:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C:Accession: AB3445
R:DeIvecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella meliten
A:Reference number: AD3252; PMID:11756688
A:Accession: AB3445
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-180 <KUR>
A:Cross-references: GB:AB008917; PIDN:AAL52725.1; PID:g17983555; GSPDB:GN00190
A:Experimental source: strain 16M
A:Genetics:
A:Gene: BMEI1544
A:Map position: 1
A:Keywords: hydrolase

```

```

Query Match      1.3%; Score 7; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 332 LKAEER 338
DB 125 LKAEER 131

```

RESULT 54

```

E70209

```

conserved hypothetical protein BBA21 - Lyme disease spirochete plasmid A/1p54
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C:Accession: E70209
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, S.D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, S.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B. Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943; PMID:9403585
 A:Accession: E70209
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-181 <KLE>
 A:Cross-references: GB:AE000790; NID:G2690224; PIDN:AAC66248.1; PID:G2690247; TIGR:BBA21
 A:Experimental source: strain B31
 C:Genetics:
 A:Genome: plasmid
 C:Superfamily: Lyme disease spirochete plasmid conserved hypothetical protein BBA21

Query Match 1.3%; Score 7; DB 2; Length 181;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 EKDFLLK 348
 DB 162 EKDFLLK 168
 |||||

RESULT 55
 E95846
 Probable ABC transporter integral membrane protein Smb20034 [imported] - Sinorhizobium meliloti
 C:Species: Sinorhizobium meliloti
 C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: E95846
 R:Pinan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan, Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: E95846
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-187 <KUR>
 A:Cross-references: GB:AL591985; PIDN:CAC48437.1; PID:gl5139909; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymB
 R:Gallbert, F.; Finan, T.W.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, P.; Barloy-Hubler, Pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T. Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, Hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K. A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: Smb20034
 A:Genome: plasmid

Query Match 1.3%; Score 7; DB 2; Length 187;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 VSEELSR 61
 DB 43 VSEELSR 49
 |||||

RESULT 56
 T45013
 Hypothetical protein [imported] - Methanosarcina acetivorans plasmid pc2A
 C:Species: Methanosarcina acetivorans
 C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000

C:Accession: T45013
 R:Metcalf, W.M.; Zhang, J.K.; Apolinario, E.; Sowers, K.R.; Wolfe, R.S. Proc. Natl. Acad. Sci. U.S.A. 94, 2626-2631, 1997
 A:Title: A genetic system for Archaea of the genus Methanosarcina: liposome-mediated tr
 A:Reference number: Z22897; MUID:97226004; PMID:9122246
 A:Accession: T45013
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-190 <MET>
 A:Cross-references: EMBL:U78295; NID:gl763609; PIDN:AAB39747.1; PID:gl763613
 A:Experimental source: strain C2A
 C:Genetics:
 A:Genome: plasmid pc2A

Query Match 1.3%; Score 7; DB 2; Length 190;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SSPQPE 8
 DB 133 SSPQPE 139
 |||||

RESULT 57
 A81149
 outer membrane lipoprotein lplB, probable NMB0873 [imported] - Neisseria meningitidis (C:Species: Neisseria meningitidis
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C:Accession: A81149
 R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A. ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M. Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A:Reference number: A81000; MUID:20175755; PMID:10710307
 A:Accession: A81149
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-193 <TET>
 A:Cross-references: GB:AE002439; GB:AE002098; NID:G7226100; PIDN:AAF41284.1; PID:G72261 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: NMB0873

Query Match 1.3%; Score 7; DB 2; Length 193;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 SELSRQL 63
 DB 102 SELSRQL 108
 |||||

RESULT 58
 H81874
 probable outer membrane lipoprotein NMA1091 [imported] - Neisseria meningitidis (strain C:Species: Neisseria meningitidis
 C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: H81874
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; More, ; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream, Nature 404, 502-506, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
 A:Reference number: A81775; MUID:20222556; PMID:10761919
 A:Accession: H81874
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-193 <PAR>
 A:Cross-references: GB:AL162755; GB:AL157959; NID:G7379742; PIDN:CA884354.1; PID:G73797 A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: NMA1091

```
Query Match      1.3%; Score 7; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 EELSRQL 63
DB 102 EELSRQL 108
|||||

RESULT 59
SAVLH1
delta large antigen - hepatitis delta virus
N/Alternate names: HDag
C/Species: hepatitis delta virus
A/Note: host Homo sapiens (man)
C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jun-2000
C/Accession: A35219
R/Saldanha, J.A.; Thomas, H.C.; Monjardino, J.P.
J. Gen. Virol. 71, 1603-1606, 1990
A/Title: Cloning and sequencing of RNA of hepatitis delta virus isolated from human serum
A/Reference number: A35219; MUID:90324949; PMID:2374010
A/Accession: A35219
A/Molecule type: genomic RNA
A/Residues: 1-195 <SAL>
A/Cross-references: GB:D01075; NID:G221689; PIDN:BA00874.1; PID:G221690
C/Comment: This virus is a replication-defective hepatitis B virus.
C/Superfamily: hepatitis delta virus large antigen
C/Keywords: core protein
F:2-195/Product: delta large antigen #status predicted <MAT>

Query Match      1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45
|||||

RESULT 60
SAVLH1
delta large antigen - hepatitis delta virus (strain Lebanon-1)
N/Alternate names: HDag
C/Species: hepatitis delta virus
A/Note: host Homo sapiens (man)
C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 01-Aug-1997
C/Accession: A40247
R/Lee, C.W.; Bih, F.Y.; Chao, Y.C.; Govindarajan, S.; Lai, M.M.C.
Virology 188, 265-273, 1992
A/Title: Evolution of hepatitis delta virus RNA during chronic infection.
A/Reference number: A40247; MUID:92230225; PMID:1566577
A/Accession: A40247
A/Molecule type: genomic RNA
A/Residues: 1-195 <LEE>
A/Cross-references: GB:M84917
C/Comment: This virus is a replication-defective hepatitis B virus.
C/Superfamily: hepatitis delta virus large antigen
C/Keywords: core protein
F:2-195/Product: delta large antigen #status predicted <MAT>

Query Match      1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45
|||||

RESULT 61
SAVLDM
delta large antigen - hepatitis delta virus (strain Japanese M-1)
```

```
N/Alternate names: HDag
C/Species: hepatitis delta virus
C/Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jun-2000
C/Accession: A36409
R/Imazeki, F.; Omata, M.; Onto, M.
J. Virol. 64, 5594-5599, 1990
A/Title: Heterogeneity and evolution rates of delta virus RNA sequences.
A/Reference number: A36409; MUID:91012805; PMID:2214027
A/Accession: A36409
A/Molecule type: genomic RNA
A/Residues: 1-195 <IMA>
A/Cross-references: GB:D90190; GB:M58299; NID:G221691; PIDN:BAAL4214.1; PID:G221692
C/Superfamily: hepatitis delta virus large antigen
F:2-195/Product: delta large antigen #status predicted <MAT>

Query Match      1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45
|||||

RESULT 62
SS3112
delta antigen - hepatitis delta virus (patient A)
C/Species: hepatitis delta virus
A/Variety: patient A
C/Date: 08-Jul-1995 #sequence_revision 01-Nov-1996 #text_change 26-Aug-1999
C/Accession: SS3112
R/Lai, M.E.; Mazzoleni, A.P.; Porru, A.; Balestrieri, A.
submitted to the EMBL Data Library, March 1995
A/Reference number: SS3112
A/Accession: SS3112
A/Molecule type: genomic RNA
A/Residues: 1-195 <LAI>
A/Cross-references: EMBL:X85253; NID:G732708; PIDN:CAAS9509.1; PID:G732709
A/Experimental source: patient A
A/Note: the source is designated as hepatitis D virus
C/Superfamily: hepatitis delta virus large antigen
C/Keywords: core protein
F:2-195/Product: delta large antigen #status predicted <MAT>

Query Match      1.3%; Score 7; DB 2; Length 195;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45
|||||

RESULT 63
G75028
h-transferring ATP synthase, chain E (atpe) PAB1182 - Pyrococcus abyssi (strain Orsay)
C/Species: Pyrococcus abyssi
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 15-Sep-2003
C/Accession: G75028
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A/Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str
A/Reference number: A75001
A/Accession: G75028
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-199 <KAW>
A/Cross-references: GB:AJ248288; GB:AL096836; NID:G5458960; PIDN:CAB50669.1; PID:G545911
A/Experimental source: strain Orsay
C/Genetics:
A/Gene: PAB1182
C/Superfamily: H+-transporting two-sector ATPase
```

```
Query Match      1.3%; Score 7; DB 2; Length 199;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 LLKEAVE 352
    |||||
Db 109 LLKEAVE 115

RESULT 64
G95852
conserved hypothetical protein [imported] - Sinorhizobium meliloti (strain 1021) magapla
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: G95852
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: G95852
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-201 <XUR>
A:CROSS-references: GB:AL5913985; PIDN:CAC48487.1; PID:G15139959; GSPDB:GN000167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.P.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lalaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMD20087
A:Genome: plasmid

Query Match      1.3%; Score 7; DB 2; Length 201;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 478 GSUTDSG 484
    |||||
Db 112 GSUTDSG 118

RESULT 65
C69375
hypothetical protein AF1003 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: C69375
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
F.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.P.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: C69375
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-211 <MLE>
A:CROSS-references: GB:AE001034; GB:AE000782; NID:G2689357; PIDN:AAB90239.1; PID:G264959

Query Match      1.3%; Score 7; DB 2; Length 211;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 TLSTPEE 171
```

```
Db 33 TLSTPEE 39
    |||||
RESULT 66
SAVLNN
delta large antigen - hepatitis delta virus (strain Nauru)
N:Alternate names: HDAG
C:Species: hepatitis delta virus
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C:Accession: A36212
R:Chao, Y.C.; Chang, M.F.; Gust, I.; Lai, M.M.C.
Virology 178, 384-392, 1990
A:Title: Sequence conservation and divergence of hepatitis delta virus RNA.
A:Reference number: A36212; MUID:91020976; PMID:2219700
A:Accession: A36212
A:Molecule type: genomic RNA
A:Residues: 1-214 <CHA>
A:CROSS-references: EMBL:IM34325; EMBL:M58629; NID:G329994; PIDN:AAB59753.1; PID:G329995
R:Otto, J.C.; Casey, P.J.
J. Biol. Chem. 271, 4569-4572, 1996
A:Title: The hepatitis delta virus large antigen is farnesylated both in vitro and in a
A:Reference number: A58234; MUID:96214864; PMID:8617711
A:Contents: annotation; confirmation of farnesylation rather than geranyl-geranylation
C:Superfamily: hepatitis delta virus large antigen
C:Keywords: core protein; lipoprotein; methylated carboxyl end; prenylated cysteine
P:2-211/Product: delta large antigen #status predicted <MAR>
P:211/Binding site: farnesyl (Cys) (covalent) #status experimental
P:211/Modified site: methyl ester carboxyl end (Cys) (in mature form) #status predicted

Query Match      1.3%; Score 7; DB 1; Length 214;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
    |||||
Db 39 KKIKKLE 45

RESULT 67
A53175
delta large antigen - hepatitis delta virus (strain Taiwan)
N:Alternate names: HDAG
C:Species: hepatitis delta virus
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 09-Jun-2000
C:Accession: A53175; PD0009
R:Chao, Y.C.; Lee, C.M.; Tang, H.S.; Govindarajan, S.; Lai, M.M.C.
Hepatology 13, 345-352, 1991
A:Title: Molecular cloning and characterization of an isolate of hepatitis delta virus :
A:Reference number: A53175; MUID:91139069; PMID:1995441
A:Accession: A53175
A>Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-214 <CHA>
A:CROSS-references: GB:M36590; NID:G329991
R:Tan, W.J.; Zhan, M.Y.; Miao, J.; Cong, X.
Chinese J. Virol. 12, 307-316, 1996
A:Title: Genetic characteristics of hepatitis delta antigens from different isolation o
A:Reference number: PD0009
A:Accession: PD0009
A:Molecule type: mRNA
A:Residues: 1-214 <TAN>
C:Superfamily: hepatitis delta virus large antigen
C:Keywords: core protein; lipoprotein; methylated carboxyl end; prenylated cysteine
P:211/Binding site: farnesyl (Cys) (covalent) #status predicted
P:211/Modified site: methyl ester carboxyl end (Cys) (in mature form) #status predicted

Query Match      1.3%; Score 7; DB 2; Length 214;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
    |||||
```



```

D5      39 KKIKKLE 45

RESULT 68
JC1062
delta large antigen - hepatitis delta virus (strain Sichuan)
N;Alternate names: HDAG
C;Species: hepatitis delta virus
C;Date: 08-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 01-Aug-1997
C;Accession: JC1062
R;Liu, S.L.; Yi, T.J.; Cong, X.; Zhan, M.Y.
Chinese J. Virol. 9, 15-22, 1993
A;Title: Molecular cloning and sequencing of a hepatitis delta antigen-coding cDNA from
A;Reference number: JC1062
A;Accession: JC1062
A;Molecule type: mRNA
A;Residues: 1-214 <LIU>
C;Superfamily: hepatitis delta virus large antigen
C;Keywords: core protein; lipoprotein; methylated carboxyl end; prenylated cysteine
F;211/Binding site: farnesyl (Cys) (covalent) #status predicted
F;211/Modified site: methyl ester carboxyl end (Cys) (in mature form) #status predicted

Query Match      1.3%; Score 7; DB 2; Length 214;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      404 KKIKKLE 410
Db      39 KKIKKLE 45

RESULT 69
S73675
hypothetical protein yjfv - Mycoplasma pneumoniae (strain ATCC 29342)
N;Alternate names: hypothetical protein P02_orf218
C;Species: Mycoplasma pneumoniae
A;Variety: ATCC 29342
C;Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
A;Accession: S73675
R;Hammelreich, R.; Hilbert, H.; Plagens, H.; Pickl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A;Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A;Reference number: S73327; MUID:97105885; PMID:8948633
A;Accession: S73675
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-218 <HLM>
A;Cross-references: EMBL:AE000033; GB:U00089; NID:G1674022; PIDN:AB95997.1; PID:G167402
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C;Genetics:
A;Gene: yjfv
A;Genetic code: SGC3
C;Superfamily: hypothetical protein H11024

Query Match      1.3%; Score 7; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      264 TLNDIQL 270
Db      175 TLNDIQL 181

RESULT 70
T05514
hypothetical protein F13M23.70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C;Accession: T05514
R;Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15419
A;Accession: T05514

A;Molecule type: DNA
A;Residues: 1-225 <BEV>
A;Cross-references: EMBL:AL035523
A;Experimental source: cultivar Columbia; BAC clone F13M23
C;Genetics:
A;Map position: 4
A;Introns: 90/2; 128/3; 152/1
A;Note: F13M23.70

Query Match      1.3%; Score 7; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      291 ERLKKLI 297
Db      205 ERLKKLI 211

RESULT 71
POBP82
antiterminal protein Q - phage 82
N;Alternate names: regulatory protein Q
C;Species: phage 82
C;Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 03-Nov-2003
C;Accession: A29791; S65585
R;Golliger, J.A.; Roberts, J.W.
J. Biol. Chem. 262, 11721-11725, 1987
A;Title: Bacteriophage 82 gene Q and Q protein: sequence, overproduction, and activity
A;Reference number: A29791; MUID:87308148; PMID:3624233
A;Accession: A29791
A;Molecule type: DNA
A;Residues: 1-229 <GOL>
A;Cross-references: GB:J02803; NID:G215364; PIDN:AAA32298.1; PID:G215365
A;Note: the authors translated the codon ACT for residue 188 as Phe and GAA for residue
R;Mahdi, A.A.; Sharples, G.J.; Mandal, T.N.; Lloyd, R.G.
J. Mol. Biol. 257, 561-573, 1996
A;Title: Holliday junction resolvases encoded by homologous rusa genes in Escherichia c
A;Reference number: S65579; MUID:96196428; PMID:8648624
A;Accession: S65585
A;Molecule type: DNA
A;Residues: 1-229 <VAH>
A;Cross-references: EMBL:CAA63332.1; PIDN:CAA63332.1; PID:G1051118
C;Genetics:
A;Gene: Q
C;Superfamily: antitermination protein Q
C;Keywords: DNA binding; late protein; transcription regulation

Query Match      1.3%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      291 ERLKKLI 297
Db      144 ERLKKLI 150

RESULT 72
D90831
antitermination protein [imported] - Escherichia coli (strain O157:H7, substrain RMD 0
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Nov-2003
C;Accession: D90831
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Fan, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90831
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA35043.1; PID:G13361084; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RMD 0509952

```

C:Genetics:

A:Gene: ECs1620

C:Superfamily: antitermination protein Q

Query Match 1.3%; Score 7; DB 2; Length 229;

Best Local Similarity 100.0%; Pred. No. 97;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 ERLKXLI 297

Db 144 ERLKXLI 150

RESULT 73

C90912

probable antitermination protein [imported] - Escherichia coli (strain O157:H7, substrain C90912)

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Nov-2003

C:Accession: C90912

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference number: A99629; PMID:11258796

A:Reference number: A99629; PMID:11258796

A:Accession: C90912

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-229 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA035690.1; PID:gl3361733; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs2267

C:Superfamily: antitermination protein Q

Query Match 1.3%; Score 7; DB 2; Length 229;

Best Local Similarity 100.0%; Pred. No. 97;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 ERLKXLI 297

Db 144 ERLKXLI 150

RESULT 74

F90972

antiterminator [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Nov-2003

C:Accession: F90972

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference number: A99629; PMID:11258796

A:Reference number: A99629; PMID:11258796

A:Accession: F90972

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-229 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA035690.1; PID:gl3362218; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs2750

C:Superfamily: antitermination protein Q

Query Match 1.3%; Score 7; DB 2; Length 229;

Best Local Similarity 100.0%; Pred. No. 97;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 ERLKXLI 297

Db 144 ERLKXLI 150

RESULT 75

G85688

probable antiterminator Q of prophage CP-933X Z1874 [imported] - Escherichia coli (stra-

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 03-Nov-2003

C:Accession: G85688

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; PMID:21074935; PMID:11206551

A:Accession: G85688

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-229 <STO>

A:Cross-references: GB:AE005174; NID:gl2514802; PIDN:AGS5971.1; GSPDB:GN00145; UWGP:Z1

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z1874

C:Superfamily: antitermination protein Q

Query Match 1.3%; Score 7; DB 2; Length 229;

Best Local Similarity 100.0%; Pred. No. 97;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 ERLKXLI 297

Db 144 ERLKXLI 150

RESULT 76

C85820

antiterminator [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 03-Nov-2003

C:Accession: C85820

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; PMID:21074935; PMID:11206551

A:Accession: C85820

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-229 <STO>

A:Cross-references: GB:AE005174; NID:gl2516127; PIDN:AGS7023.1; GSPDB:GN00145; UWGP:Z3;

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z3114

C:Superfamily: antitermination protein Q

Query Match 1.3%; Score 7; DB 2; Length 229;

Best Local Similarity 100.0%; Pred. No. 97;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 ERLKXLI 297

Db 144 ERLKXLI 150

RESULT 77

G90292

hypothetical protein SS01363 [imported] - Sulfolobus solfataricus transposon ISC048

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: G90292

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, I.

arratt, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: G90292

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-232 <KUR>
A:Cross-references: GB:AE006641; NID:gl3814574; PIDN:AAK41598.1; GSPDB:GN00155
C:Genetics:
A:Gene: SSO1363

Query Match 1.3%; Score 7; DB 2; Length 232;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 404 KKIKLE 410
Db 53 KKIKLE 59
|||||

RESULT 78
E97347
probable transcription regulator, containing DNA-binding domain of xre family [imported]
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: E97347
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium CLO
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: E97347
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-232 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK81568.1; PID:gl5026747; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3646

Query Match 1.3%; Score 7; DB 2; Length 232;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 322 DAKLQQA 328
Db 163 DAKLQQA 169
|||||

RESULT 79
A86150
TIN6.24 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C:Accession: A86150
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chin, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.; Hansen, N.P.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A86150
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-235 <STO>
A:Cross-references: GB:AE005172; NID:g8671852; PIDN:AAF78415.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 1.3%; Score 7; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 406 IKKLEKE 412
Db 152 IKKLEKE 158
|||||

RESULT 80

H97687
Hypothetical 17.5K protein in avtA-selB intergenic region (ol57a) [imported] - Agrobact
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: H97687
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Iappas, C.; Markelt, B.
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H97687
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-237 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK88457.1; PID:gl5157960; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_4972
A:Map position: circular chromosome

Query Match 1.3%; Score 7; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 VSEELSR 61
Db 104 VSEELSR 110
|||||

RESULT 81

B95374
Hypothetical protein Sma1644 [imported] - Sinorhizobium meliloti (strain 1021) magaplas
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: B95374
R:Barnett, M.J.; Fisher, R.P.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bow
A.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilo
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: B95374
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-239 <KUR>
A:Cross-references: GB:AS006469; PIDN:AAK65556.1; PID:gl4524034; GSPDB:GN00165
A:Experimental source: Strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, A.; Ampe, P.; Barloy-Hubler
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.P.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure
neault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Sma1644
A:Genome: plasmid

Query Match 1.3%; Score 7; DB 2; Length 239;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 215 HSKAVLA 221
Db 52 HSKAVLA 58
|||||

RESULT 82

F90885
hypothetical protein ECs2054 [imported] - Escherichia coli (strain O157:H7, substrain R1)
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90885
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome
A:Reference number: A99829; MUID:21156231; PMID:11258796
A:Accession: F90885
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-240 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA835477.1; PID:gl13361520; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 050952
C:Genetics:
A:Gene: ECs2054

Query Match 1.3%; Score 7; DB 2; Length 240;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 KLOQAQE 330
Db 137 KLOQAQE 143

RESULT 83

A85733
hypothetical protein Z2269 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85733
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-240 <STO>
A:Cross-references: GB:AE005174; NID:gl12515249; PIDN:AAG56325.1; GSPDB:GN00145; UWGP:222
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2269

Query Match 1.3%; Score 7; DB 2; Length 240;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 KLOQAQE 330
Db 137 KLOQAQE 143

RESULT 84

E64897
hypothetical protein b1450 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: E64897
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E64897
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-240 <BLAT>

A:Cross-references: GB:AB000242; GB:U00096; NID:gl1787720; PIDN:AACT4532.1; PID:gl1787722
A:Experimental source: strain K-12, substrain MG1655

Query Match 1.3%; Score 7; DB 2; Length 240;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 KLOQAQE 330
Db 137 KLOQAQE 143

RESULT 85

H72732
hypothetical protein AB0401 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 18-Aug-2003
C:Accession: H72732
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Taka
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: H72732
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-246 <KAW>
A:Cross-references: DDBJ:AP000059; NID:gs103911; PIDN:BAA79356.1; PID:gs104040
A:Experimental source: strain K1
C:Genetics:
A:Gene: AB0401
C:Superfamily: uridylylate kinase

Query Match 1.3%; Score 7; DB 2; Length 246;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 ROLEDIL 67
Db 186 ROLEDIL 192

RESULT 86

JH0252
myelin P0 protein precursor - human
N:Alternate names: myelin protein zero
C:Species: Homo sapiens (man)
C>Date: 31-Mar-1992 #sequence_revision 06-Dec-1996 #text_change 16-Jun-2000
R:Hayasaka, K.; Nanao, K.; Tahara, M.; Sato, W.; Takada, G.; Miura, M.; Uyemura, K.
Biochem. Biophys. Res. Commun. 180, 515-518, 1991
A:Title: Isolation and sequence determination of cDNA encoding the major structural pro
A:Reference number: JH0252; MUID:92062068; PMID:1719967
A:Accession: JH0252
A:Molecule type: mRNA
A:Residues: 1-248 <HAY>
A:Cross-references: GB:D10537; GB:D90501; NID:g220073; PIDN:BAA01395.1; PID:g220074
A:Experimental source: fetus spinal cord
R:Hayasaka, K.; Ohnishi, A.; Takada, G.; Fukushima, Y.; Murai, Y.
Biochem. Biophys. Res. Commun. 194, 1317-1322, 1993
A:Title: Mutation of the myelin P0 gene in charcot-marie-tooth neuropathy type 1.
A:Reference number: JH0704; MUID:93356807; PMID:7688964
A:Accession: JH0704
A:Molecule type: mRNA
A:Residues: 1-248 <HAY>
R:Pham-Dinh, D.; Fourbil, Y.; Blanquet, F.; Mattei, M.G.; Roeckel, N.; Latour, P.; Chaz
Rum. Mol. Genet. 2, 2051-2054, 1993
A:Title: The major peripheral myelin protein zero gene: structure and localization in tl
A:Reference number: 139378; MUID:94154677; PMID:7509228
A:Accession: 139378
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

```

submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15385
A:Accession: T04809
A:Molecule type: DNA
A:Residues: 1-262 <BV>
A:Cross-references: EMBL:AL035440
A:Experimental source: cultivar Columbia; BAC clone F10M23
C:Genetics:
A:Map position: 4
A:A1 introns: 74/3; 97/1; 112/3; 142/1; 210/2; 237/3
A>Note: F10M23-200
C:Superfamily: conserved hypothetical protein H10090

Query Match          1.3%; Score 7; DB 2; Length 262;
Best Local Similarity 100.0%; Pred.No. 1.1e+02;
Matches      7; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

QY      21 AAPAVEA 27
DB       2 MAPAVEA 8
|||||

RESULT 89
T37222
probable secreted protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T37222
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21615
A:Accession: T37222
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-271 <OLI>
A:Cross-references: EMBL:AL031514; PIDN:CAA20600.1; GSPDB:GN00070; SCOEDB:SC2H4.06c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC2H4.06c

Query Match          1.3%; Score 7; DB 2; Length 271;
Best Local Similarity 100.0%; Pred.No. 1.1e+02;
Matches      7; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

QY      41 GAQARTA 47
DB     113 GAQARTA 119
|||||

RESULT 90
JC6142
deoxyguanosine kinase (EC 2.7.1.113) precursor - human
C:Species: Homo sapiens (man)
C>Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 18-Jun-1999
C:Accession: JC6142
R:Johansson, M.; Karlsson, A.
Proc. Natl. Acad. Sci. U.S.A. 93, 7258-7262, 1996
A>Title: Cloning and expression of human deoxyguanosine kinase cDNA.
A>Note: it is uncertain whether Met-1 or Met-8 is the initiator
A:Reference number: JC6142; MUID:96293511; PMID:8692979
A:Accession: JC6142
A:Molecule type: mRNA
A:Residues: 1-277 <JOH>
A:Cross-references: GS:T041668; NID:g1477481; PIDN:AAC50624.1; PID:g1477482
A>Note: It is uncertain whether Met-1 or Met-8 is the initiator
C:Comment: This enzyme is involved in mediating cytotoxicity of nucleoside analogs.
C:Keywords: thymidine kinase 2.
C:Superfamily: human deoxycytidine kinase
C:Keywords: phosphotransferase
F.I-1-39/Domain: signal sequence #status predicted <SIG>
P.F:40-277/Product: deoxyguanosine kinase #status predicted <MAT>

Query Match          1.3%; Score 7; DB 2; Length 277;
Best Local Similarity 100.0%; Pred.No. 1.2e+02;

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 QPAREEE 251
 Db 205 QPAREEE 211

RESULT 91
 A:Title: carboxylesterase (RC 3.1.1.1) [imported] - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
 C:Accession: A10182
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarrage, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Li, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: A10182
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-280 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CA030324.1; PID:gl5979543; GSPDB:GN00175
 C:Genetics:
 A:Gene: YPO1501
 C:Superfamily: conserved hypothetical protein YJL068c
 C:Keywords: carboxylic ester hydrolase

Query Match 1.3%; Score 7; DB 2; Length 280;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 ELLEHR 188
 Db 6 ELLEHR 12

RESULT 92
 C86423
 A:Title: unknown protein, 14666-16092 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: C86423
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, K.; Hughes, B.; Huizlar, L.; Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: AB6141; MUID:21016719; PMID:11130712
 A:Accession: C86423
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-285 <STO>
 A:Cross-references: GB:AB005172; NID:gl0092456; PIDN:AGL12858.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 1.3%; Score 7; DB 2; Length 285;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 SDEVGR 138
 Db 44 SDEVGR 50

RESULT 93

F64750
 A:Title: regulatory protein perR - Escherichia coli (strain K-12)
 N:Alternate names: Peroxidase resistance protein perR
 C:Species: Escherichia coli
 C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: F64750
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: F64750
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-297 <BLAT>
 A:Cross-references: GB:AE000133; GB:U00096; NID:G2367099; PIDN:AA073357.1; PID:gl786448
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: perR
 C:Superfamily: regulatory protein ampR
 C:Keywords: DNA binding; transcription regulation
 F:23-53/Region: regulatory protein lysR motif
 F:221-237/Domain: transmembrane #status predicted <TM>

Query Match 1.3%; Score 7; DB 2; Length 297;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 LAERLKK 295
 Db 173 LAERLKK 179

RESULT 94
 T21938
 A:Title: hypothetical protein F38A3.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2003
 C:Accession: T21938
 R:Swinburne, J.
 submitted to the EMBL Data Library, June 1995
 A:Reference number: Z19490
 A:Accession: T21938
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-306 <WIL>
 A:Cross-references: EMBL:Z49938; PIDN:CAA90187.1; GSPDB:GN00020; CESP:F38A3.1
 A:Experimental source: Clone F38A3
 C:Genetics:
 A:Gene: CESP:F38A3.1
 A:Map position: 2
 A:Introns: 26/2; 54/3

Query Match 1.3%; Score 7; DB 2; Length 306;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 GPGEDG 83
 Db 277 GPGEDG 283

RESULT 95
 G84701
 A:Title: hypothetical protein At2g29880 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: G84701
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, E. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G34701
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-308 <STO>
A;Cross-references: GB:AE002093; NID:g4567213; PIDN:AAD23628.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g29880
A;Map position: 2

Query Match 1.3%; Score 7; DB 2; Length 308;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 336 EERHORE 342
Db 242 EERHORE 248

RESULT 96

D64303
hypothetical protein H11305 homolog - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 07-Jul-2003
C;Accession: D64303
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. Science 273, 1058-1073, 1996
A;Authors: Kane, B.F.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C. A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: D64303
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-319 <BUL>
A;Cross-references: GS:U67461; GB:I77117; NID:GL590827; PIDN:ABE98008.1; PID:GL590830; C;Genetics:
C;Map position: REV29473-28514
C;Superfamily: thiamine monophosphate kinase

Query Match 1.3%; Score 7; DB 2; Length 319;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 GLGKEIT 157
Db 222 GLGKEIT 228

RESULT 97

T45568
hypothetical protein Fl1C1.60 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T45568
R;Bargues, M.; Collado, M.C.; Navarro, P.; Terol, J.; Perez-Alonso, M.; Mewes, H.W.; May submitted to the Protein Sequence Database, December 1999
A;Reference number: 223007
A;Accession: T45568
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-322 <BAR>
A;Cross-references: EMBL:AL132976
A;Experimental source: cultivar Columbia; BAC clone Fl1C1
C;Genetics:
A;Map position: 3
A;Note: Fl1C1.60

Query Match 1.3%; Score 7; DB 2; Length 322;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 AVLARSK 224
Db 249 AVLARSK 255

RESULT 98

E64547
nitrogen fixation protein nifU homolog - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: E64547
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C. A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori. A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: E64547
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-326 <TOM>
A;Cross-references: GB:AE000542; GB:AE000511; NID:G2313310; PIDN:AAD07289.1; PID:G23133 C;Superfamily: Helicobacter nitrogen fixation protein; nitrogen fixation protein homoio F;53-106/Domain: nitrogen fixation protein homology <NPH>

Query Match 1.3%; Score 7; DB 1; Length 326;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 465 LNKRVQD 471
Db 96 LNKRVQD 102

RESULT 99

B71960
hypothetical protein jhp0207 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Jul-1999
C;Accession: B71960
R;Alm, R.A.; Ling, R.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jfang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 387, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric pat. A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: B71960
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-326 <ARN>
A;Cross-references: GB:AE001459; GB:AE001439; NID:G4154723; PIDN:AAD05790.1; PID:G41547 A;Experimental source: strain J99
C;Genetics:

A;Gene: jhp0207
C;Superfamily: Helicobacter nitrogen fixation protein; nitrogen fixation protein homolo F;53-106/Domain: nitrogen fixation protein homology <NPH>

Query Match 1.3%; Score 7; DB 2; Length 326;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 465 LNKRVQD 471
Db 96 LNKRVQD 102

RESULT 100

B82918
DNA-directed RNA polymerase alpha chain U0257 [imported] - Ureaplasma urealyticum
C;Species: Ureaplasma urealyticum
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: B82918

R:Glass, J.I.; Zeffkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
 submitted to GenBank, February 2000
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mir
 A:Reference number: A82870
 A:Accession: B82918
 A>Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-327 <GLA>
 A:Cross-references: GB:AE002123; GB:AF222894; MID:g6899229; PIDN:AAF30666.1; GSPDB:GN001
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:Gene: rpoA; U0257
 A:Genetic code: SGC3
 C:Superfamily: DNA-directed RNA polymerase alpha chain

Query Match 1.3%; Score 7; DB 2; Length 327;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 REEEEXR 254
 |||||
 DB 254 REEEEXR 260

Search completed: June 7, 2004, 14:46:55
 Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:38:11 ; Search time 48 Seconds
(without alignments)
3119.795 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 530

Sequence: 1 KSPGPGPRAGPEGAQRPSQ.....APSTASGQTGPQPTPSARA 530

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DS seq length: 0

Maximum DS seq length: 2000000000

Post-processing: Listing first 300 summaries

Database : A Geneseq 29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	530	100.0	530	2	AAW49042 Human low
2	530	100.0	530	4	AAB82804 Human low
3	429	80.9	546	4	AAB82808 Human low
4	429	80.9	639	6	ABR41210 Human DIT
5	429	80.9	639	6	ABR41186 Human DIT
6	359	67.7	510	4	ABR11764 Human LDL
7	359	67.7	510	4	AAW79741 Human pro
8	356	67.2	356	6	ABO07223 Human p53
9	347	65.5	358	4	AAW78757 Human pro
10	228	43.0	1749	4	ABG00839 Novel hum
11	172	32.5	557	2	AAW49039 Rabbit lo
12	172	32.5	557	4	AAW49039 Rabbit lo
13	19	3.6	204	4	AAW83281 Human bre
14	19	3.6	386	2	AAW37883 BRCAL mod
15	19	3.6	386	2	AAW30151 Amiro aci
16	19	3.6	395	6	ABO07222 Human p53
17	19	3.6	528	4	AAB93594 Human pro
18	17	3.2	85	5	ABP42281 Human ova
19	17	3.2	676	2	AAV17863 Neurite e
20	17	3.2	684	6	ABR43284 Human neu
21	12	2.3	455	4	AAO86301 Human pol
22	12	2.3	475	7	ADC33216 Human nov
23	9	1.7	99	4	AAU22308 Human car
24	9	1.7	99	7	ADB46276 Human car
25	9	1.7	554	4	ABBS9454 Drosophil

26	9	1.7	637	6	ABU44191	Abu44191 Protein e
27	9	1.7	729	2	AAW89275	AAW89275 Yeast coa
28	8	1.5	65	2	AAV12665	AAV12665 Human 5'
29	8	1.5	156	2	AAV01195	AAV01195 Polypteti
30	8	1.5	170	3	AAW08727	AAW08727 Arabidops
31	8	1.5	174	3	AAW08726	AAW08726 Arabidops
32	8	1.5	194	4	ABG19843	ABG19843 Novel hum
33	8	1.5	224	3	AAW20570	AAW20570 Arabidops
34	8	1.5	229	5	ABU05665	ABU05665 M. tuberc
35	8	1.5	360	5	ABG77243	ABG77243 Selected
36	8	1.5	360	5	ABJ11114	ABJ11114 Yeast sel
37	8	1.5	364	5	ABJ11057	ABJ11057 Yeast sel
38	8	1.5	388	7	ADB95026	ADB95026 A. thalia
39	8	1.5	416	7	ADC31613	ADC31613 Human nov
40	8	1.5	484	4	AAW79356	AAW79356 Corynebact
41	8	1.5	484	4	AAW92676	AAW92676 C glutami
42	8	1.5	785	3	AAW41279	AAW41279 Arabidops
43	8	1.5	823	3	AAW41278	AAW41278 Arabidops
44	8	1.5	836	3	AAW41277	AAW41277 Arabidops
45	8	1.5	952	7	ADB95084	ADB95084 A. thalia
46	8	1.5	1045	4	ABG08332	ABG08332 Novel hum
47	8	1.5	1532	3	AAW40945	AAW40945 Human ORF
48	7	1.3	7	4	AAW82820	AAW82820 Human low
49	7	1.3	15	5	AAW50054	AAW50054 Prolin-e
50	7	1.3	15	7	ADD24103	ADD24103 Breast ca
51	7	1.3	18	6	ABP83398	ABP83398 G protein
52	7	1.3	25	2	AAW03444	AAW03444 HDV anti
53	7	1.3	27	3	AAW64949	AAW64949 Human 5'
54	7	1.3	32	2	AAW03445	AAW03445 HDV anti
55	7	1.3	34	6	ABP97458	ABP97458 HIV N36-b
56	7	1.3	34	6	ABP97457	ABP97457 HIV N36-b
57	7	1.3	35	2	AAW03447	AAW03447 HDV anti
58	7	1.3	35	2	AAW03446	AAW03446 HDV anti
59	7	1.3	36	4	ABG23756	ABG23756 Novel hum
60	7	1.3	36	4	ABG27966	ABG27966 Novel hum
61	7	1.3	36	4	ABG28172	ABG28172 Novel hum
62	7	1.3	37	2	AAW03443	AAW03443 HDV anti
63	7	1.3	37	4	ABG27959	ABG27959 Novel hum
64	7	1.3	38	2	AAW03442	AAW03442 HDV anti
65	7	1.3	38	4	AAW16323	AAW16323 Peptide #
66	7	1.3	38	4	ABR35315	ABR35315 Peptide #
67	7	1.3	38	4	AAW28818	AAW28818 Peptide #
68	7	1.3	38	4	ABR30145	ABR30145 Peptide #
69	7	1.3	38	4	ABE20757	ABE20757 Protein bon
70	7	1.3	38	4	AAW68519	AAW68519 Human bon
71	7	1.3	38	4	AAW56146	AAW56146 Human liv
72	7	1.3	38	4	ABG50188	ABG50188 Human liv
73	7	1.3	38	4	AAW04061	AAW04061 Peptide #
74	7	1.3	38	5	ABG38100	ABG38100 Human pep
75	7	1.3	42	4	ABG28175	ABG28175 Novel hum
76	7	1.3	43	4	ABG23760	ABG23760 Novel hum
77	7	1.3	43	4	ABG23761	ABG23761 Novel hum
78	7	1.3	44	3	AAW95964	AAW95964 Construct
79	7	1.3	45	3	AAV76347	AAV76347 Fragment
80	7	1.3	45	3	ABW56241	ABW56241 Human sec
81	7	1.3	45	3	ADE11987	ADE11987 Human sec
82	7	1.3	50	2	AAW03431	AAW03431 HDV anti
83	7	1.3	50	2	AAW03428	AAW03428 Delta hep
84	7	1.3	50	2	AAW03438	AAW03438 HDV anti
85	7	1.3	50	2	AAW03439	AAW03439 HDV anti
86	7	1.3	50	2	AAW03437	AAW03437 HDV anti
87	7	1.3	50	2	AAW03436	AAW03436 HDV anti
88	7	1.3	50	2	AAW03434	AAW03434 HDV anti
89	7	1.3	50	2	AAW03440	AAW03440 HDV anti
90	7	1.3	51	4	ABG10316	ABG10316 Novel hum
91	7	1.3	52	4	ABG23757	ABG23757 Novel hum
92	7	1.3	52	4	AAW03441	AAW03441 HDV anti
93	7	1.3	57	2	ABG27957	ABG27957 Novel hum
94	7	1.3	60	7	ADC00659	ADC00659 Enterohae
95	7	1.3	70	4	ABG20222	ABG20222 Novel hum
96	7	1.3	72	4	ABG27972	ABG27972 Novel hum
97	7	1.3	73	4	AAO01466	AAO01466 Human pol

99	7	1.3	75	4	ABG25987	Abg25987 Novel hum	172	7	1.3	156	4	ABG08702	Abg08702 Novel hum
100	7	1.3	75	4	ABG28171	Abg28171 Novel hum	173	7	1.3	156	5	Au82951	Au82951 Ruman hom
101	7	1.3	75	4	ABG27974	Abg27974 Novel hum	174	7	1.3	158	4	ABG08703	Abg08703 Novel hum
102	7	1.3	76	4	ABG25983	Abg25983 Novel hum	175	7	1.3	159	6	ABU06063	Abu06063 Human mat
103	7	1.3	76	4	ABG16198	Abg16198 Novel hum	176	7	1.3	169	6	ABU70381	Abu70381 Human adi
104	7	1.3	76	5	ABP00042	Abp00042 Human ORF	177	7	1.3	173	2	AY07084	AY07084 Renal can
105	7	1.3	78	4	AAU59856	Aau59856 Propionib	178	7	1.3	173	4	ABG08700	Abg08700 Novel hum
106	7	1.3	78	6	ABM52375	Abm52375 Propionib	179	7	1.3	173	4	ABG02018	Abg02018 Novel hum
107	7	1.3	80	4	AAU44520	Aau44520 Propionib	180	7	1.3	174	4	ABG02018	Abg02018 Novel hum
108	7	1.3	80	6	ABM43039	Abm43039 Propionib	181	7	1.3	174	4	ABG30142	Abg30142 Novel hum
109	7	1.3	82	4	AAU80866	Aau80866 Human hae	182	7	1.3	174	4	ABG20594	Abg20594 Novel hum
110	7	1.3	85	3	AAU01980	Aau01980 Human sec	183	7	1.3	176	2	AY25609	AY25609 Parietari
111	7	1.3	85	4	ABG27963	Abg27963 Novel hum	184	7	1.3	176	2	ADC34859	Adc34859 Parietari
112	7	1.3	85	4	ABG09671	Abg09671 Novel hum	185	7	1.3	177	5	ABP66324	Abp66324 Bifidobac
113	7	1.3	85	4	ABG28298	Abg28298 Novel hum	186	7	1.3	186	4	ABP12097	Abp12097 Human neu
114	7	1.3	86	5	ABP98832	Abp98832 Human pro	187	7	1.3	187	6	ADA35994	Ada35994 Acinetoba
115	7	1.3	87	4	ABG23765	Abg23765 Novel hum	188	7	1.3	188	4	ABG11317	Abg11317 Human AFP
116	7	1.3	91	4	AAU81329	Aau81329 Human hae	189	7	1.3	190	4	ABG16195	Abg16195 Novel hum
117	7	1.3	91	4	AAU80872	Aau80872 Human hae	190	7	1.3	193	3	AY75683	AY75683 Neisseria
118	7	1.3	91	4	ABG09698	Abg09698 Novel hum	191	7	1.3	193	3	AY75682	AY75682 Neisseria
119	7	1.3	95	4	ABP24335	Abp24335 Protein #	192	7	1.3	193	4	ABG27973	Abg27973 Novel hum
120	7	1.3	95	4	ABG54726	Abg54726 Human liv	193	7	1.3	194	2	ABG22383	Abg22383 Antigen t
121	7	1.3	95	4	ABG09669	Abg09669 Novel hum	194	7	1.3	194	7	ABG29071	Abg29071 Human hgx
122	7	1.3	95	5	ABG42854	Abg42854 Human pep	195	7	1.3	197	3	AY79205	AY79205 Soluble h
123	7	1.3	98	4	ABG10187	Abg10187 Novel hum	196	7	1.3	197	6	ABU60682	Abu60682 Human sol
124	7	1.3	98	4	ABG22793	Abg22793 Novel hum	197	7	1.3	201	4	ABG03304	Abg03304 Novel hum
125	7	1.3	98	4	ABG10185	Abg10185 Novel hum	198	7	1.3	201	4	ABG27969	Abg27969 Novel hum
126	7	1.3	98	4	ABG00647	Abg00647 Novel hum	199	7	1.3	201	4	ABG23585	Abg23585 Novel hum
127	7	1.3	101	4	ABG28176	Abg28176 Novel hum	200	7	1.3	201	4	ABG23762	Abg23762 Novel hum
128	7	1.3	102	4	ABG27960	Abg27960 Novel hum	201	7	1.3	204	4	AAU38884	Aau38884 Human pol
129	7	1.3	102	4	ABG27626	Abg27626 Novel hum	202	7	1.3	205	4	AAU38883	Aau38883 Human pol
130	7	1.3	104	4	ABG01408	Abg01408 Novel hum	203	7	1.3	207	7	ADB64453	ADB64453 Human pro
131	7	1.3	105	4	ADC88181	Adc88181 Ribosomal	204	7	1.3	208	4	ABG23759	Abg23759 Novel hum
132	7	1.3	106	5	ABP43161	Abp43161 Human ova	205	7	1.3	208	4	ABG12942	Abg12942 Novel hum
133	7	1.3	114	4	ABG01409	Abg01409 Novel hum	206	7	1.3	211	5	ABP40330	Abp40330 Staphyloc
134	7	1.3	115	4	ABG27958	Abg27958 Novel hum	207	7	1.3	213	3	AGS53748	AGS53748 Arabidops
135	7	1.3	115	5	ABG70160	Abg70160 Human pre	208	7	1.3	213	3	AGS53773	AGS53773 Arabidops
136	7	1.3	117	2	AAW55515	Aaw55515 H. pylori	209	7	1.3	213	3	AGI16923	AGI16923 Arabidops
137	7	1.3	117	4	AAU01194	Aau01194 Polypepti	210	7	1.3	213	3	AGI16923	AGI16923 Arabidops
138	7	1.3	118	4	AAU01194	Aau01194 Polypepti	211	7	1.3	213	3	AGI16923	AGI16923 Arabidops
139	7	1.3	118	4	AAU00818	Abg00818 Novel hum	212	7	1.3	213	3	AGI16923	AGI16923 Arabidops
140	7	1.3	118	4	ABG18560	Abg18560 Novel hum	213	7	1.3	213	3	AGI16923	AGI16923 Arabidops
141	7	1.3	118	4	ABG27964	Abg27964 Novel hum	214	7	1.3	215	3	AGI16923	AGI16923 Arabidops
142	7	1.3	118	4	ABG00452	Abg00452 Novel hum	215	7	1.3	215	3	ABU56245	Abu56245 Human sec
143	7	1.3	118	4	ABG27627	Abg27627 Novel hum	216	7	1.3	215	6	ABU70766	Abu70766 Human adi
144	7	1.3	122	2	AAU01194	Aau01194 Polypepti	217	7	1.3	215	6	ABU70766	Abu70766 Human adi
145	7	1.3	127	4	AAO02070	Aao02070 Human pol	218	7	1.3	216	6	ABU70648	Abu70648 Human adi
146	7	1.3	127	7	ADE61282	Ad61282 Rat Prote	219	7	1.3	218	3	AGS53772	AGS53772 Arabidops
147	7	1.3	129	4	ABG08365	Abg08365 Novel hum	220	7	1.3	218	3	AGI16922	AGI16922 Arabidops
148	7	1.3	130	3	AGS24902	AGS24902 Arabidops	221	7	1.3	218	3	AGS53747	AGS53747 Arabidops
149	7	1.3	133	2	AAW55583	Aaw55583 H. pylori	222	7	1.3	219	3	AGI18069	AGI18069 Arabidops
150	7	1.3	133	2	AAW55697	Aaw55697 H. pylori	223	7	1.3	220	4	ABP36582	Abp36582 Human FLE
151	7	1.3	133	2	AAW55584	Aaw55584 H. pylori	224	7	1.3	220	5	AAO21540	Aao21540 Human rib
152	7	1.3	133	2	AAU25607	Aau25607 Parietari	225	7	1.3	225	4	ABG26532	Abg26532 Novel hum
153	7	1.3	133	7	ADC34857	Adc34857 Parietari	226	7	1.3	228	4	ABG10401	Abg10401 Novel hum
154	7	1.3	137	2	AAW55385	Aaw55385 H. pylori	227	7	1.3	228	6	ABU20130	Abu20130 Protein e
155	7	1.3	138	2	AAU25610	Aau25610 Parietari	228	7	1.3	229	7	ADC00597	Adc00597 Enteroha
156	7	1.3	138	7	ADC34860	Adc34860 Parietari	229	7	1.3	229	7	ADC00070	Adc00070 Enteroha
157	7	1.3	139	5	AAU79009	Aau79009 P. judaic	230	7	1.3	229	7	ADC00747	Adc00747 Enteroha
158	7	1.3	139	5	AAU79012	Aau79012 P. judaic	231	7	1.3	234	4	ABG26550	Abg26550 Novel hum
159	7	1.3	139	5	AAU79008	Aau79008 P. judaic	232	7	1.3	234	4	ABG27965	Abg27965 Novel hum
160	7	1.3	139	5	AAU79011	Aau79011 P. judaic	233	7	1.3	234	4	ABG22794	Abg22794 Novel hum
161	7	1.3	139	5	AAU79010	Aau79010 P. judaic	234	7	1.3	235	3	AAU24748	Aau24748 Plant SDF
162	7	1.3	140	7	ADE31106	Ad31106 Human dia	235	7	1.3	238	6	ABU21072	Abu21072 Protein e
163	7	1.3	142	5	ABG50085	Abg50085 Human DIF	236	7	1.3	240	6	ABU15384	Abu15384 Protein e
164	7	1.3	143	2	AAU25611	Aau25611 Parietari	237	7	1.3	240	6	ABP2263	Abp2263 GPBP-inte
165	7	1.3	143	7	ADC34861	Adc34861 Parietari	238	7	1.3	240	6	ABP2263	Abp2263 GPBP-inte
166	7	1.3	146	3	AAU24901	Aau24901 Arabidops	239	7	1.3	241	4	ABG10855	Abg10855 Novel hum
167	7	1.3	146	4	ABG20929	Abg20929 Novel hum	240	7	1.3	244	4	AAU40669	Aau40669 Human pol
168	7	1.3	150	4	ABG30011	Abg30011 Novel hum	241	7	1.3	244	4	AAU40670	Aau40670 Human pol
169	7	1.3	153	4	ABG25410	Abg25410 Novel hum	242	7	1.3	245	4	ABG03146	Abg03146 Novel hum
170	7	1.3	155	6	ABU02638	Abu02638 S. pneumo	243	7	1.3	251	4	ABG08208	Abg08208 Novel hum
171	7	1.3	156	4	ABG01403	Abg01403 Novel hum	244	7	1.3	251	7	ADB31909	ADB31909 Plant (A.

245 7 1.3 258 5 AAM52696 Human mye
 246 7 1.3 261 2 AAW70493 Human dis
 247 7 1.3 262 4 AAG19805 Novel hum
 248 7 1.3 261 4 ABG19805 Novel hum
 249 7 1.3 264 6 ABG64618 Drosophil
 250 7 1.3 264 6 ABU31369 Protein e
 251 7 1.3 265 6 ABG62248 GPBP-inte
 252 7 1.3 266 5 ABG62248
 253 7 1.3 269 4 ABG19807
 254 7 1.3 271 5 AAU96713 Human ost
 255 7 1.3 275 6 ABG27796 Protein e
 256 7 1.3 276 4 ABG60199 Drosophil
 257 7 1.3 276 4 ABG64125 Human pro
 258 7 1.3 277 5 AAG75049 Human col
 259 7 1.3 277 5 AAE16384 Human geo
 260 7 1.3 277 6 ABU31840 Protein e
 261 7 1.3 280 6 ABU50303 Protein e
 262 7 1.3 282 4 ABG25339 Novel hum
 263 7 1.3 282 5 AAU96712 Human ost
 264 7 1.3 289 5 ABP59081 Renal cel
 265 7 1.3 293 3 AAG37986 Human sec
 266 7 1.3 293 7 ADC87089 Human GPC
 267 7 1.3 295 3 AAG18068 Arabidops
 268 7 1.3 296 4 AAG68186 PINCH-lik
 269 7 1.3 296 6 ABR41110 Human PIN
 270 7 1.3 296 7 ADR98730 PINCH-lik
 271 7 1.3 296 7 ADR82526 Human pro
 272 7 1.3 297 4 ABG29977 Novel hum
 273 7 1.3 302 4 AAU31152 Novel hum
 274 7 1.3 302 6 ADA33221 Acinetoba
 275 7 1.3 305 3 AAG18067 Arabidops
 276 7 1.3 305 4 ABG10404 Novel hum
 277 7 1.3 309 2 AAY28246 Human sec
 278 7 1.3 309 4 AAU39023 Human sec
 279 7 1.3 309 5 ABB55732 Human pol
 280 7 1.3 310 6 ABR58027 Rat GI en
 281 7 1.3 310 7 ADB5641 Human pro
 282 7 1.3 312 2 AAR27361 Sequence
 283 7 1.3 317 6 ABU25194 Protein e
 284 7 1.3 320 6 ABW71403 Staphyloc
 285 7 1.3 322 3 AAG53771 Arabidops
 286 7 1.3 322 3 AAG56921 Arabidops
 287 7 1.3 322 3 AAG25639 Arabidops
 288 7 1.3 326 2 AAG53746 Arabidops
 289 7 1.3 326 4 AAU35892 Helicobac
 290 7 1.3 326 4 ABG00257 Novel hum
 291 7 1.3 326 6 ABU30728 Protein e
 292 7 1.3 327 4 AAU31453 Novel hum
 293 7 1.3 327 6 ABU48864 Protein e
 294 7 1.3 327 7 ADD71119 Human int
 295 7 1.3 329 4 AAU20543 Human sec
 296 7 1.3 335 4 AAB94947 Human pro
 297 7 1.3 336 6 AAO16355 Human IRS
 298 7 1.3 337 3 AAB16695 Bacteriop
 299 7 1.3 337 5 ABB04329 Human sec
 300 7 1.3 339 3 AAG53382 Arabidops

ALIGNMENTS

RESULT 1
 AAW49042
 ID AAW49042 standard; protein; 530 AA.
 XX
 AC AAW49042;
 XX
 DT 09-NOV-1998 (first entry)
 XX Human low density lipoprotein binding protein LBP-3.
 DE
 XX Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;
 XX receptor; human; atherosclerosis; diagnosis; therapy; vaccine.
 KW

XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 XX Peptide 53..59
 FT /note= "Claim 2"
 FT
 XX
 PW WO9823282-A1.
 XX
 PD 04-JUN-1998.
 XX
 PF 26-NOV-1997; 97WO-US021857.
 XX
 PR 27-NOV-1996; 96US-0031930P.
 PR 03-JUN-1997; 97US-0048547P.
 XX
 PA (BOST-) BOSTON HEART FOUND INC.
 XX
 PI Lees AM, Lees RS, Law SW, Arjona AA;
 DR WPI; 1998-322455/28.
 DR N-P8DB; AAV32839.
 XX
 PT Nucleic acid encoding low density lipoprotein binding proteins and
 PT related vectors - transformed cells, proteins, and modulators of binding,
 PT useful for treatment and diagnosis of atherosclerosis and for identifying
 PT subjects at risk.
 XX
 PS Claim 1; Fig 8; 47pp; English.
 XX
 CC This polypeptide comprises novel human low density lipoprotein (LDL)
 CC binding protein LBP-3 that is capable of binding both native and methyl
 CC LDL. Its amino acid sequence was deduced from an isolated cDNA clone (see
 CC AAV32839). cDNA clones (see AAV32834-39) and encoded rabbit and human
 CC LBPs (see AAW49037-42) are claimed. An abnormality in an aspect of LBP
 CC metabolism or structure is diagnostic of a risk for atherosclerosis. The
 CC invention provides methods for determining if an animal is at risk for
 CC atherosclerosis (e.g. for prenatal screening); methods for treating
 CC atherosclerosis (including gene therapy) using e.g. LBP polypeptides to
 CC bind LDL and thereby prevent formation of atherosclerotic plaque; and
 CC methods for treating a cell having an abnormality in LBP structure or
 CC metabolism. Pharmaceutical and vaccine compositions are also provided, as
 CC well as recombinant vectors and host cells used to produce recombinant
 CC LBP
 XX
 SQ Sequence 530 AA;
 XX
 Query Match 100.0%; Score 530; DB 2; Length 530;
 Best-Local Similarity 100.0%; Pred. No. 0;
 Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KSSPQPEAGPEGAGQERPSQAAPAVEARGPSSQAAPRKPGEQAQARTAGSGLRDVSEILS 60
 DB 1 KSSPQPEAGPEGAGQERPSQAAPAVEARGPSSQAAPRKPGEQAQARTAGSGLRDVSEILS 60
 QY 61 RQLEDILSTYCVDNNGGPGEDGAQGPAPEDAEKSRITYVARNGEPEPTPVYGEKEPS 120
 DB 61 RQLEDILSTYCVDNNGGPGEDGAQGPAPEDAEKSRITYVARNGEPEPTPVYGEKEPS 120
 QY 121 KGDPTNTEIRQSDVGDHRRPQEKKAGLGEKXITLLMOTLNTLSTPEEKLAALCKKY 180
 DB 121 KGDPTNTEIRQSDVGDHRRPQEKKAGLGEKXITLLMOTLNTLSTPEEKLAALCKKY 180
 QY 181 AELLLEHNSQKMKLQKKOSQLVQEKDHLRGHSKAVLARSKLESICRELQNRSLK 240
 DB 181 AELLLEHNSQKMKLQKKOSQLVQEKDHLRGHSKAVLARSKLESICRELQNRSLK 240
 QY 241 EGVQRAEEEEKKREKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLIQY 300
 DB 241 EGVQRAEEEEKKREKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLIQY 300
 QY 301 ELREHIDKVPKHQLOQLOQVDAKLOQAEMLKAEERHOREKDFLLKEAVESQKCELM 360

Db 301 ELREEHIDKVPKHKDLQQLVDAKIQQAQEWLKEAERHOREKDFLLKEAVESQRMCELM 360
 QY 361 KOETHLKOQALALYTEKFEFQNTLSKSSVFPTTFQEMEKMTKKIKKLEKETTMYRSRW 420
 Db 361 KOETHLKOQALALYTEKFEFQNTLSKSSVFPTTFQEMEKMTKKIKKLEKETTMYRSRW 420
 QY 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVDLSAGQGGSL 480
 Db 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVDLSAGQGGSL 480
 QY 481 TDGSPRRPBGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQEPSTARA 530
 Db 481 TDGSPRRPBGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQEPSTARA 530
 RESULT 2
 AAB82804
 ID AAB82804 standard; protein; 530 AA.
 AC AAB82804;
 XX
 DT 12-NOV-2001 (first entry)
 DE Human low density lipoprotein binding protein 3 (LBP-3).
 DE Low density lipoprotein binding protein 3; LBP-3; LDL; human;
 KW atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.
 XX OS Homo sapiens.
 XX WO200164874-A2.
 XX 07-SEP-2001.
 XX 28-FEB-2001; 2001WO-US006356.
 XX 02-MAR-2000; 2000US-00517849.
 PR 14-JUL-2000; 2000US-00616289.
 XX (BOST-) BOSTON HEART FOUND INC.
 FA Lees AM, Lees RS, Law SW, Arjona AA;
 XX WPI; 2001-565505/63.
 XX N-PSDB; AAH26496.
 XX New isolated low density lipoprotein binding polypeptide for treating,
 PT diagnosing and/or identifying therapeutic agents for atherosclerosis.
 XX Claim 13(h); Fig 8B; 143pp; English.
 XX The present sequence is that of the N-terminal portion of novel human low
 CC density lipoprotein binding protein 3 (LBP-3). The amino acid sequence is
 CC deduced from an isolated partial cDNA clone (see AAH26496). A full-length
 CC sequence is given in AAB82804. Human LBP-3 is an example of claimed LBP
 CC polypeptides of the invention that are capable of binding to native and
 CC methylated low density lipoproteins. Also claimed are biologically active
 CC fragments and analogues of LBPs, polynucleotides encoding LBPs, as well
 CC as expression vectors, cells, and methods of producing the LBPs. Methods
 CC of determining if an animal is at risk for atherosclerosis, methods for
 CC evaluating an agent for use in treating atherosclerosis, and methods for
 CC treating a cell having an abnormality in structure or metabolism of LBP
 CC are claimed. Pharmaceutical compositions comprising an LBP polypeptide or
 CC nucleic acid, and vaccine compositions, are also claimed
 XX Sequence 530 AA;
 SQ
 Query Match 100.0%; Score 530; DB 4; Length 530;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KSSPGQPEAGPEGAQRPSPQAAPAVEAGPGSQAPRKPEGAQARTAQSGALRDVSEELS 60
 |||||

Db 1 KSSPGQPEAGPEGAQRPSPQAAPAVEAGPGSQAPRKPEGAQARTAQSGALRDVSEELS 60
 QY 61 ROLEDILSTYCVNNKQGGEDGAGGPAEDADAESKSTYVARNGEPPEPTPVVYGEKPS 120
 Db 61 ROLEDILSTYCVNNKQGGEDGAGGPAEDADAESKSTYVARNGEPPEPTPVVYGEKPS 120
 QY 121 KGDPNTEBIROSDEVDGRDHRPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 180
 Db 121 KGDPNTEBIROSDEVDGRDHRPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 180
 QY 181 ABLLBEHNSQVKKLLOKQSQOLVOEKDHLRGEHSKAVLARSKLESICRELORNRSLK 240
 Db 181 ABLLBEHNSQVKKLLOKQSQOLVOEKDHLRGEHSKAVLARSKLESICRELORNRSLK 240
 QY 241 EESVQARAEEREEKKEVTSHPOVTINDIQLQWEOHNRNSKLFQENMELAEELKKLIEQY 300
 Db 241 EESVQARAEEREEKKEVTSHPOVTINDIQLQWEOHNRNSKLFQENMELAEELKKLIEQY 300
 QY 301 ELREEHIDKVPKHKDLQQLVDAKIQQAQEWLKEAERHOREKDFLLKEAVESQRMCELM 360
 Db 301 ELREEHIDKVPKHKDLQQLVDAKIQQAQEWLKEAERHOREKDFLLKEAVESQRMCELM 360
 QY 361 KOETHLKOQALALYTEKFEFQNTLSKSSVFPTTFQEMEKMTKKIKKLEKETTMYRSRW 420
 Db 361 KOETHLKOQALALYTEKFEFQNTLSKSSVFPTTFQEMEKMTKKIKKLEKETTMYRSRW 420
 QY 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVDLSAGQGGSL 480
 Db 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVDLSAGQGGSL 480
 QY 481 TDGSPRRPBGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQEPSTARA 530
 Db 481 TDGSPRRPBGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQEPSTARA 530
 RESULT 3
 AAB82808
 ID AAB82808 standard; protein; 546 AA.
 AC AAB82808;
 XX
 DT 12-NOV-2001 (first entry)
 DE Human low density lipoprotein binding protein 3 (LBP-3).
 DE Low density lipoprotein binding protein 3; LBP-3; LDL; human;
 KW atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.
 XX OS Homo sapiens.
 XX WO200164874-A2.
 XX 07-SEP-2001.
 XX 28-FEB-2001; 2001WO-US006356.
 XX 02-MAR-2000; 2000US-00517849.
 PR 14-JUL-2000; 2000US-00616289.
 XX (BOST-) BOSTON HEART FOUND INC.
 FA Lees AM, Lees RS, Law SW, Arjona AA;
 XX WPI; 2001-565505/63.
 XX N-PSDB; AAH26501.
 XX New isolated low density lipoprotein binding polypeptide for treating,
 PT diagnosing and/or identifying therapeutic agents for atherosclerosis.
 XX Claim 13(l); Fig 8A; 143pp; English.
 XX The present sequence is that of novel human low density lipoprotein
 CC binding protein 3 (LBP-3). The amino acid sequence was deduced from the

CC coding region of isolated genomic DNA (see AB26501). It differs from the
CC sequence predicted from an LBP-3 cDNA clone (see AB2804) by the
CC presence of an additional 16 amino acids at the N-terminus (the cDNA
CC clone is 5' truncated) and by having asparagine at amino acid position
CC 130 rather than tyrosine. Human LBP-2 is an example of claimed LBP
CC polypeptides of the invention that are capable of binding to native and
CC methylated low density lipoproteins. Also claimed are biologically active
CC fragments and analogues of these LBPs, polynucleotides encoding LBPs, as
CC well as expression vectors, cells and methods of producing the LBPs. A
CC polypeptide having the amino acid residues 96-110 of the present sequence
CC is claimed (see AB2820). Methods of determining if an animal is at risk
CC for atherosclerosis, methods for evaluating an agent for use in treating
CC atherosclerosis, and methods for treating a cell having an abnormality in
CC structure or metabolism of LBP are claimed. Pharmaceutical compositions
CC comprising an LBP polypeptide or nucleic acid, and vaccine compositions,
CC are also claimed
XX

SQ Sequence 546 AA;

Query Match 80.9%; Score 429; DB 4; Length 546;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	KSPGQPEAGPGEAQRPSQAPAVEAEGPGSQAPRKPGEQAQRTAGSGLRDVSESL	60
DB	17	KSPGQPEAGPGEAQRPSQAPAVEAEGPGSQAPRKPGEQAQRTAGSGLRDVSESL	76
QY	61	ROLEDLITVYVDNNOGGGEGDGAQGEPAEPDAEKSRVTVARNGEPTPTVVVGEKSPS	120
DB	77	ROLEDLITVYVDNNOGGGEGDGAQGEPAEPDAEKSRVTVARNGEPTPTVVVGEKSPS	136
QY	121	KGDPNTEETQSDVEGDRDHRPQEKKAGKGLGKITLLMOTLNTLSTPEEKLAALCKKY	180
DB	137	KGDPNTEETQSDVEGDRDHRPQEKKAGKGLGKITLLMOTLNTLSTPEEKLAALCKKY	196
QY	181	AZLLHEHNSQOMKLLQKQSQOLVQEKDHLRGEHSKAVLARSKLESICRELQNRSLK	240
DB	197	AZLLHEHNSQOMKLLQKQSQOLVQEKDHLRGEHSKAVLARSKLESICRELQNRSLK	256
QY	241	EGVQARAREEKRKEVTSHPQVTLNDIQLOMEQHNERNSKLRQENMELAEKLLIROY	300
DB	257	EGVQARAREEKRKEVTSHPQVTLNDIQLOMEQHNERNSKLRQENMELAEKLLIROY	316
QY	301	ELREEHIDKVFHKDLQOOLVDAKLOQAEMLKEAEERHOREKDFLLAEVDSQMCML	360
DB	317	ELREEHIDKVFHKDLQOOLVDAKLOQAEMLKEAEERHOREKDFLLAEVDSQMCML	376
QY	361	KQOETHLQOALYATEKFEFQVTLKSSEVPTTPKQEMKMTKIKKLEKETTMYRSRW	420
DB	377	KQOETHLQOALYATEKFEFQVTLKSSEVPTTPKQEMKMTKIKKLEKETTMYRSRW	436
QY	421	ESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERNDLNRVQDLSAGGQSL	480
DB	437	ESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERNDLNRVQDLSAGGQSL	496
QY	481	TDSPGPRRPGGAQAPSPRVTEAPCPVGPAPSTASGQTGPQEPPTSARA	530
DB	497	TDSPGPRRPGGAQAPSPRVTEAPCPVGPAPSTASGQTGPQEPPTSARA	546

RESULT 4

ABR41210

ID ABR41210 standard; protein; 639 AA.

XX

AC ABR41210;

XX

DT 02-JUN-2003 (first entry)

DE Human DITHP extracellular signalling protein.

XX

KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;

KW cancer; cell proliferative disorder; autoimmune disorder;

KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;

KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW extracellular signalling.
OS Homo sapiens.
XI
FN WO200297031-A2.
XX
PD 05-DEC-2002.
XX
PF 27-MAR-2002; 2002WO-US010056.
XX
PR 28-MAR-2001; 2001US-0279619P.
PR 29-MAR-2001; 2001US-0280067P.
PR 29-MAR-2001; 2001US-0280068P.
PR 16-MAY-2001; 2001US-0291280P.
PR 17-MAY-2001; 2001US-0291829P.
PR 17-MAY-2001; 2001US-0291849P.
PR 19-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-0300001P.
XX
PA (INCY-) INCYTE GENOMICS INC.
PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amehey SR;
PI Daugherty SC, Dam TC, Liu TP, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX WPI; 2003-129518/12.
DR N-PSDB; ACC46153.
XX
PT Novel human diagnostic and therapeutic polypeptide useful for identifying
PT test compound which specifically binds to a polypeptide encoded by human
PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX
PS Claim 27; SEQ ID NO 745; 591bp; English.
XX
CC The invention relates to novel human diagnostic and therapeutic
CC polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC proteins (DITHP; ABR4136-ABR41812). The invention also relates to
CC polynucleotide sequences at least 90% identical to the dithp cDNA
CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dithp nucleic acid sequence; the
CC recombinant production of DITHP proteins; antibodies specific for DITHP
CC proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC detecting dithp nucleotide and protein sequences; methods of screening
CC for compounds which specifically bind a DITHP protein; and methods of
CC assessing the toxicity of test compounds using a dithp hybridisation
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC diagnosis of a wide variety of conditions including cancer and other cell
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC viral, fungal or parasitic infections; hormonal disorders; metabolic
CC disorders; neurological disorders; gastrointestinal disorders; transport
CC disorders; and connective tissue disorders. They may also be used to
CC screen for modulators of protein activity or gene expression. DITHP
CC proteins can additionally be used in analysis of the proteome of a tissue
CC or cell type and to induce antibodies. The dithp nucleic acids are
CC additionally useful in somatic or germline gene therapy of the disorders
CC mentioned above, as a source of antisense sequences, as a source of
CC probes and primers, in genotyping and identification of individuals, in
CC the generation of transgenic animal models of human disease or knock in
CC humanised animals, in toxicological testing, and in transcript imaging.
CC The present sequence represents a DITHP protein which has extracellular
CC signalling activity. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from Wipo at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 639 AA;

Query Match 80.9%; Score 429; DB 6; Length 639;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPSSQAAPRKEGAQARTAQSGALRDVSEELS 60
DB 110 KSSPGQPEAGPEGAQERPSQAAPAVEAGPSSQAAPRKEGAQARTAQSGALRDVSEELS 169

QY 61 RQLEDILSTYCVNNQGGFGDGAQGEPAEPEDAEKSRITYVARNGEPEFTPVVYGEKEPS 120
DB 170 RQLEDILSTYCVNNQGGFGDGAQGEPAEPEDAEKSRITYVARNGEPEFTPVVYGEKEPS 229

QY 121 KGDPTNTEIRQSDVEGDRDRHRRPQKKAAGLGEITLLMQTLNTLSTPEEKLAALCKY 180
DB 230 KGDPTNTEIRQSDVEGDRDRHRRPQKKAAGLGEITLLMQTLNTLSTPEEKLAALCKY 289

QY 181 AELLEHRNSQOMKLLQKKQSLQVQKHRLRGEHSKAVLARSKLESICRELQRHNSLK 240
DB 290 AELLEHRNSQOMKLLQKKQSLQVQKHRLRGEHSKAVLARSKLESICRELQRHNSLK 349

QY 241 EGVQRAEEBEKKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 300
DB 350 EGVQRAEEBEKKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 409

QY 301 ELREHIDKVPKHDLQOOLVDKLLQQAQEMLKAEERHOREKDFLLKEAVESORMCELM 360
DB 410 ELREHIDKVPKHDLQOOLVDKLLQQAQEMLKAEERHOREKDFLLKEAVESORMCELM 469

QY 361 KQCEHLKQALALYTKPEFBFQNTLSKSSVFTTFKQEMEWTKKIKKLEKETTMYRSRW 420
DB 470 KQCEHLKQALALYTKPEFBFQNTLSKSSVFTTFKQEMEWTKKIKKLEKETTMYRSRW 529

QY 421 ESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERNDLNKEVODLSAGGQSL 480
DB 530 ESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERNDLNKEVODLSAGGQSL 589

QY 481 TDSGPRPEGAQAPSPRVTAPCPYGPAPSTASQQTGPQPTSARA 530
DB 590 TDSGPRPEGAQAPSPRVTAPCPYGPAPSTASQQTGPQPTSARA 639

RESULT 5
ABR41186
ID ABR41186 standard; protein; 639 AA.
XX
AC ABR41186;
XX
DT 02-JUN-2003 (first entry)
XX
DE Human DITHP extracellular signalling protein.
XX
KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW extracellular signalling.
XX
OS Homo sapiens.
XX
FN WO2000297031-A2.
XX
PD 05-DEC-2002.
XX
PF 27-MAR-2002; 2002WO-US010056.
XX
PR 28-MAR-2001; 2001US-0279619P.
XX
PR 29-MAR-2001; 2001US-0280047P.
XX
PR 29-MAR-2001; 2001US-0280048P.
XX
PR 16-MAY-2001; 2001US-0291280P.
XX
PR 17-MAY-2001; 2001US-0291829P.

PR 17-MAY-2001; 2001US-0291849P.
PR 19-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-0300001P.
XX (INCY-) INCYTE GENOMICS INC.
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GS, Hillman JT, Yu JY, Tuason O, Yap PE, Ansley SR;
PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX WPI; 2003-129518/12.
DR N-PSDB; ACC46130.
XX
PT Novel human diagnostic and therapeutic polypeptide useful for identifying
PT test compound which specifically binds to a polypeptide encoded by human
PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX
PS Claim 27; SEQ ID NO 721; 591pp; English.
XX
CC The invention relates to novel human diagnostic and therapeutic
CC polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC polynucleotide sequences at least 90% identical to the dithp cDNA
CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dithp nucleic acid sequence; the
CC recombinant production of DITHP proteins; antibodies specific for DITHP
CC proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC detecting dithp nucleotide and protein sequences; methods of screening
CC for compounds which specifically bind a DITHP protein; and methods of
CC assessing the toxicity of test compounds using a dithp hybridisation
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC diagnosis of a wide variety of conditions including cancer and other cell
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC viral, fungal or parasitic infections; hormonal disorders; metabolic
CC disorders; neurological disorders; gastrointestinal disorders; transport
CC disorders; and connective tissue disorders. They may also be used to
CC screen for modulators of protein activity or gene expression. DITHP
CC proteins can additionally be used in analysis of the proteome of a tissue
CC or cell type and to induce antibodies. The dithp nucleic acids are
CC additionally useful in somatic or germline gene therapy of the disorders
CC mentioned above, as a source of antisense sequences, as a source of
CC probes and primers, in genotyping and identification of individuals, in
CC the generation of transgenic animal models of human disease or knock in
CC humanised animals, in toxicological testing, and in transcript imaging.
CC The present sequence represents a DITHP protein which has extracellular
CC signalling activity. Note: the sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 639 AA;
Query Match 80.9%; Score 429; DB 6; Length 639;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPSSQAAPRKEGAQARTAQSGALRDVSEELS 60
DB 110 KSSPGQPEAGPEGAQERPSQAAPAVEAGPSSQAAPRKEGAQARTAQSGALRDVSEELS 169

QY 61 RQLEDILSTYCVNNQGGFGDGAQGEPAEPEDAEKSRITYVARNGEPEFTPVVYGEKEPS 120
DB 170 RQLEDILSTYCVNNQGGFGDGAQGEPAEPEDAEKSRITYVARNGEPEFTPVVYGEKEPS 229

QY 121 KGDPTNTEIRQSDVEGDRDRHRRPQKKAAGLGEITLLMQTLNTLSTPEEKLAALCKY 180
DB 230 KGDPTNTEIRQSDVEGDRDRHRRPQKKAAGLGEITLLMQTLNTLSTPEEKLAALCKY 289

QY 181 AELLEHRNSQOMKLLQKKQSLQVQKHRLRGEHSKAVLARSKLESICRELQRHNSLK 240
DB 290 AELLEHRNSQOMKLLQKKQSLQVQKHRLRGEHSKAVLARSKLESICRELQRHNSLK 349

CC activity, and detecting a test agent-biased activity of the assay system.
CC Also included are modulating (M2) a p53 pathway of a cell (comprising
CC contacting a cell defective in p53 function with a candidate modulator
CC that specifically binds to a HM polypeptide comprising an HM amino acid
CC sequence, where p53 function is restored), modulating (M3) a p53 pathway
CC in a mammalian cell (comprising contacting the cell with an agent that
CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
CC a disease in a patient (comprising: (a) obtaining a biological sample
CC from the patient; (b) contacting the sample with a probe for HM
CC expression; (c) comparing the results with a control; and (d) determining
CC whether the comparison indicates a likelihood disease). (M1) is useful
CC for identifying modulators of the p53 pathway. A probe for HM expression
CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
CC in a patient, where the cancer has greater than 25 % expression level.
CC Modulators identified by (M1) are useful in a variety of diagnostic and
CC therapeutic applications, where disease or disorder prognosis is related
CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
CC proliferation disorders (e.g. cancer). Another two new methods (M2 and
CC M3) are useful for modulating the p53 pathway of a cell, thus restoring
CC the p53 function of the cell, so that the cell undergoes normal
CC proliferation or progression through the cell cycle. (M2) and (M3) are
CC also useful for treating defects in the p53 pathway such as angiogenic,
CC apoptotic or cell proliferation disorders. The present sequence
CC represents a human p53 pathway modifying protein
XX
SQ Sequence 356 AA;

Query Match 67.2%; Score 356; DB 6; Length 356;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 356; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 175 ALCKYAELEEHNSQKMLQKQSLVQSKDHLRGHSAVLARSKLSLCRELOR 234
DB 1 ALCKYAELEEHNSQKMLQKQSLVQSKDHLRGHSAVLARSKLSLCRELOR 60
QY 235 HNSLKEGVQVAREEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLK 294
DB 61 HNSLKEGVQVAREEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLK 120
QY 295 KLEQYELREHIDKVFKHDLQQLVDKLLQQAQEMLKEABERHOREKDFLLKEAVESQ 354
DB 121 KLEQYELREHIDKVFKHDLQQLVDKLLQQAQEMLKEABERHOREKDFLLKEAVESQ 180
QY 355 RMCELMKQSTHLKQALYTERPEEFQNTLSKSSSEVFTTFKQEMKVTNKKIKLEKETT 414
DB 181 RMCELMKQSTHLKQALYTERPEEFQNTLSKSSSEVFTTFKQEMKVTNKKIKLEKETT 240
QY 415 MYSRWESSNKALLEAEKTVDRDKELEGQVQIKQLEKLCRALQTERNDLNKRVQDLSA 474
DB 241 MYSRWESSNKALLEAEKTVDRDKELEGQVQIKQLEKLCRALQTERNDLNKRVQDLSA 300
QY 475 GGQGSLLTDSGPERRPGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 530
DB 301 GGQGSLLTDSGPERRPGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 356

RESULT 9
AA078757
ID AA078757 standard; protein; 358 AA.
AC AA078757;
XX
DT 06-NOV-2001 (first entry)
XX Human protein SEQ ID NO 1419.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
OS Homo sapiens.
XX

PN W0200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren P, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
DR N-PSDB; AAK51890.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX in diagnosis and gene therapy.
XX
XX Claim 20; Page 3681-3682; 5221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AA078323-AA080302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AA080020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 358 AA;

Query Match 65.5%; Score 347; DB 4; Length 358;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 184 LEEHNSQKMLQKQSLVQSKDHLRGHSAVLARSKLSLCRELQNRSLKBERG 243
DB 12 LEEHNSQKMLQKQSLVQSKDHLRGHSAVLARSKLSLCRELQNRSLKBERG 71
QY 244 VORAREEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIQYELR 303
DB 72 VORAREEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIQYELR 131
QY 304 EEHIDKVPFKHDLQQLVDKLLQQAQEMLKEABERHOREKDFLLKEAVESQRMCELMKQ 363
DB 132 EEHIDKVPFKHDLQQLVDKLLQQAQEMLKEABERHOREKDFLLKEAVESQRMCELMKQ 191
QY 364 ETHLKQALYTERPEEFQNTLSKSSSEVFTTFKQEMKVTNKKIKLEKETTMYSRWESS 423
DB 192 ETHLKQALYTERPEEFQNTLSKSSSEVFTTFKQEMKVTNKKIKLEKETTMYSRWESS 251
QY 424 NKALLEAEKTVDRDKELEGQVQIKQLEKLCRALQTERNDLNKRVQDLSAGGQSLTDS 483
DB 252 NKALLEAEKTVDRDKELEGQVQIKQLEKLCRALQTERNDLNKRVQDLSAGGQSLTDS 311
QY 484 GPERPBGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 530
DB 312 GPERPBGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 358

```
RESULT 10
ABG00839
ID ABG00839 standard; protein; 1749 AA.
XX
AC ABG00839;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #830.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
DR N-PSDB; AAS65026.
XX
WPI; 2001-639362/73.
XX
DR N-PSDB; AAS65026.
XX
New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 31198; 103pp; English.
XX
The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (III) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG0377 represent novel human diagnostic
CC patent did not appear in the invention. Note: The sequence data for this
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1749 AA;
XX
Query Match 43.0%; Score 228; DB 4; Length 1749;
Best Local Similarity 100.0%; Pred. No. 7.4e-211;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 241 EGVQVAREERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELARLKLIEQY 300
DB 438 EGVQVAREERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELARLKLIEQY 497
QY 301 ELREEHIDKVPKHKDLQQLVDAKLAQAQEMLKEABERHQKDFLLKEAVESQRMCELM 360
|||||
```

```
498 ELREEHIDKVPKHKDLQQLVDAKLAQAQEMLKEABERHQKDFLLKEAVESQRMCELM 557
361 KQOETHLKKQALALYTKFKFSEFQNTLSKSEVFTTFKQEMEMTKTKIKLSEKTTMTYRSRW 420
|||||
558 KQOETHLKKQALALYTKFKFSEFQNTLSKSEVFTTFKQEMEMTKTKIKLSEKTTMTYRSRW 617
|||||
421 ESNKALLEKAEKTVRDKELEGLQVKIQRLSEKLCALQTERNDLNKR 468
|||||
618 ESNKALLEKAEKTVRDKELEGLQVKIQRLSEKLCALQTERNDLNKR 665

RESULT 11
AAW49039
ID AAW49039 standard; protein; 557 AA.
XX
AC AAW49039;
XX
DT 09-NOV-1998 (first entry)
XX
DE Rabbit low density lipoprotein binding protein LBP-3.
XX
KW Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;
KW receptor; rabbit; atherosclerosis; diagnosis; therapy; vaccine.
XX
OS Oryctolagus cuniculus.
XX
XX
XX Key Location/Qualifiers
XX Peptide 96..110
XX /note= "Claim 2"
XX
XX WO9823282-A1.
XX
XX 04-JUN-1998.
XX
XX 26-NOV-1997; 97WO-US021857.
XX
XX 27-NOV-1996; 96US-0031930P.
XX
XX 03-JUN-1997; 97US-0048547P.
XX
XX (BOST-) BOSTON HEART FOUND INC.
XX
XX Lees AM, Lees RS, Law SW, Arjona AA;
XX
XX WPI; 1998-322455/28.
XX
XX N-PSDB; AAV32836.
XX
XX Nucleic acid encoding low density lipoprotein binding proteins and
XX related vectors - transformed cells, proteins, and modulators of binding,
XX useful for treatment and diagnosis of atherosclerosis and for identifying
XX subjects at risk.
XX
XX Claim 1; Fig 5; 47pp; English.
XX
XX This polypeptide comprises novel rabbit low density lipoprotein (LDL)
XX binding protein LBP-3 that is capable of binding both native and methyl
XX LDL. Its amino acid sequence was deduced from rabbit abdominal aorta cDNA
XX (see AAV32836). cDNA clones (see AAV32834-39) and encoded rabbit and human
XX LBPs (see AAW49037-42) are claimed. An abnormality in an aspect of LBP
XX metabolism or structure is diagnostic of a risk for atherosclerosis. The
XX invention provides methods for determining if an animal is at risk for
XX atherosclerosis (e.g. for prenatal screening); methods for treating
XX atherosclerosis (including gene therapy) using e.g. LBP polypeptides to
XX bind LDL, and thereby prevent formation of atherosclerotic plaque; and
XX methods for treating a cell having an abnormality in LBP structure or
XX metabolism. Pharmaceutical and vaccine compositions are also provided, as
XX well as recombinant vectors and host cells used to produce recombinant
XX LBP
XX
XX Sequence 557 AA;
XX
Query Match 32.5%; Score 172; DB 2; Length 557;
Best Local Similarity 100.0%; Pred. No. 4e-157;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 264 TLNDIQLQWEOHNRNSKLRQENNELAERLKKLIEQYELREHIDKVFHKDLQOOLVDA 323
DB 280 TLNDIQLQWEOHNRNSKLRQENNELAERLKKLIEQYELREHIDKVFHKDLQOOLVDA 339
QY 324 KLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKKOOLALYTEKEEFQ 383
DB 340 KLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKKOOLALYTEKEEFQ 399
QY 384 TLKSSEVFTTFKQEMERKTKIKKLEKETTMYRSRWESSNKALLEMAEKT 435
DB 400 TLKSSEVFTTFKQEMERKTKIKKLEKETTMYRSRWESSNKALLEMAEKT 451

RESULT 12
ID AAB82801 standard; protein; 557 AA.
XX
AC AAB82801;
XX
DT 12-NOV-2001 (first entry)
XX
DE Rabbit low density lipoprotein binding protein 3 (LBP-3).
XX
KW Low density lipoprotein binding protein 3; LBP-3; LDL; rabbit;
KW atherosclerosis; antiatherosclerotic; therapy; diagnosis; vaccine.
XX
OS Oryctolagus cuniculus.
XX
PN WO200164874-A2.
XX
PD 07-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US006356.
XX
PR 02-MAR-2000; 2000US-00517849.
XX
PR 14-JUL-2000; 2000US-00616289.
XX
PA (BOST-) BOSTON HEART FOUND INC.
XX
PI Lees AM, Lees RS, Law SW, Arjona AA;
XX
DR WPI; 2001-565505/63.
XX
DR N-PSDB; AAH26491.
XX
PT New isolated low density lipoprotein binding polypeptide for treating,
PT diagnosing and/or identifying therapeutic agents for atherosclerosis.
XX
PS Claim 13(e); Fig 5; 143pp; English.
XX
CC The present sequence is that of novel rabbit low density lipoprotein
CC binding protein 1 (LBP-3). The amino acid sequence is deduced from an
CC isolated cDNA clone (see AAH26491). Rabbit LBP-3 is an example of claimed
CC polypeptides of the invention, termed LBPs, that are capable of binding
CC to native and methylated low density lipoproteins. Also claimed are
CC biologically active fragments and analogues of LBPs, polynucleotides
CC encoding LBPs, as well as expression vectors, cells and methods of
CC producing the LBPs. A polypeptide having the amino acid residues 96-110
CC of the present sequence is claimed (see AAB82819). Methods of determining
CC if an animal is at risk for atherosclerosis, methods for evaluating an
CC agent for use in treating atherosclerosis, and methods for treating a
CC cell having an abnormality in structure or metabolism of LBP are claimed.
CC Pharmaceutical compositions comprising an LBP polypeptide or nucleic
CC acid, and vaccine compositions, are also claimed
XX
SQ Sequence 557 AA;

Query Match 32.5%; Score 172; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 4e-157;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 TLNDIQLQWEOHNRNSKLRQENNELAERLKKLIEQYELREHIDKVFHKDLQOOLVDA 323
DB 280 TLNDIQLQWEOHNRNSKLRQENNELAERLKKLIEQYELREHIDKVFHKDLQOOLVDA 339
QY 324 KLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKKOOLALYTEKEEFQ 383
DB 340 KLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKKOOLALYTEKEEFQ 399
QY 384 TLKSSEVFTTFKQEMERKTKIKKLEKETTMYRSRWESSNKALLEMAEKT 435
DB 400 TLKSSEVFTTFKQEMERKTKIKKLEKETTMYRSRWESSNKALLEMAEKT 451
```

```
DB 280 TLNDIQLQWEOHNRNSKLRQENNELAERLKKLIEQYELREHIDKVFHKDLQOOLVDA 339
QY 324 KLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKKOOLALYTEKEEFQ 383
DB 340 KLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKKOOLALYTEKEEFQ 399
QY 384 TLKSSEVFTTFKQEMERKTKIKKLEKETTMYRSRWESSNKALLEMAEKT 435
DB 400 TLKSSEVFTTFKQEMERKTKIKKLEKETTMYRSRWESSNKALLEMAEKT 451

RESULT 13
ID AAB63259 standard; protein; 204 AA.
XX
AC AAB63259;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human breast cancer associated antigen protein sequence SEQ ID NO:621.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine.
XX
OS Homo sapiens.
XX
PN WO200073801-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US014749.
XX
PR 28-MAY-1999; 99US-0136526P.
XX
PR 10-SEP-1999; 99US-0153454P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Obata Y;
XX
DR WPI; 2001-025274/03.
XX
PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer.
XX
PS Example 1; Page 486-487; 799pp; English.
XX
CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer
XX
SQ Sequence 204 AA;

Query Match 3.6%; Score 19; DB 4; Length 204;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 LNTLSTPEEKLAALCKKYA 181
DB 1 LNTLSTPEEKLAALCKKYA 19

RESULT 14
ID AAB37883 standard; protein; 386 AA.
```

XX AAW37883;
 AC
 XX
 DT 28-AUG-1998 (first entry)
 DE
 XX BRCA1 modulator protein 091-132Q20.
 XX
 KW BRCA1 modulator protein; 091-132Q20; breast cancer antigen 1;
 KW tumour suppressor protein; diagnosis; therapy; human.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 124..143
 FT /note="leucine zipper motif"
 XX
 XX WO9810066-A1.
 PN
 XX
 XX 12-MAR-1998.
 PD
 XX
 XX 06-AUG-1997; 97WO-US013944.
 PF
 XX
 XX 04-SEP-1996; 96US-0025601P.
 PR
 XX (ONVX-) ONVX PHARM INC.
 PA
 XX Rubinfeld B, Polakis P, Ligenfelter C, Vuong TT;
 PI
 XX WPI; 1998-193616/17.
 DR N-PSDB; AAV29064.
 DR
 XX Breast cancer antigen 1 modulator protein - useful for diagnosing
 XX diseases involving unwanted cell growth, e.g. breast cancer, and for
 XX producing therapeutics for treatment of such diseases.
 XX
 XX Example 1; Fig 3; 73pp; English.
 PS
 XX This polypeptide comprises a 46 kDa BRCA1 modulator protein that binds to
 CC the tumour suppressor gene product BRC1, and which is characterised by a
 CC leucine zipper motif. Its amino acid sequence was deduced from the
 CC nucleotide sequence of a cDNA clone (see AAV29064), designated 091-132Q20
 CC (ATCC 98143), isolated from a HeLa cell cDNA library using a yeast two-
 CC hybrid assay. 3 cDNA clones (see also AAV29062 and AAV29063) coding for
 CC BRCA1 modulator proteins (see AAW37881-83) have been characterised.
 CC Vectors and host cells comprising the isolated nucleic acid sequences are
 CC claimed, as well as a process for producing BRCA1 modulator protein by
 CC culturing these host cells. BRCA1 modulator proteins and nucleic acids
 CC can be used to diagnose diseases involving unwanted cell growth, e.g.
 CC breast cancer, and to identify compounds that alter BRCA1 interaction
 CC with BRCA1 modulators for the treatment of such diseases
 XX
 SQ Sequence 386 AA;
 Query Match 3.6%; Score 19; DB 2; Length 386;
 Best Local Similarity 100.0%; Pred. No. 2.5e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 163 LNTLSTPEKLAALCKKYA 181
 DB 135 LNTLSTPEKLAALCKKYA 153
 RESULT 15
 ID AAY30151 standard; protein; 386 AA.
 AC AAY30151;
 XX
 XX 27-OCT-1999 (first entry)
 DT
 XX Amino acid sequence of a BRCA1 modulator protein.
 DE
 XX Modulator protein; BRCA1; tumour suppressor protein; breast cancer;
 KW

KW ovarian cancer; cell growth; cell proliferation.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 125..143
 FT /note="leucine zipper motif"
 XX
 XX US5948643-A.
 PN
 XX
 XX 07-SEP-1999.
 PD
 XX
 XX 13-AUG-1997; 97US-00968751.
 PF
 XX
 XX 13-AUG-1997; 97US-00968751.
 PR
 XX (ONVX-) ONVX PHARM INC.
 PA
 XX Rubinfeld B, Ligenfelter C, Vuong TT, Polakis PG;
 PI
 XX WPI; 1999-517952/43.
 DR N-PSDB; AAX86756.
 DR
 XX Modulator proteins that bind to and modulate the activity of the BRCA1
 XX tumor suppressor gene product, useful for the treatment of ovarian and
 XX breast cancer.
 XX
 XX Example 1; Fig 3; 35pp; English.
 PS
 XX The present sequence represents a modulator protein, that binds to and
 CC modulate the activity of the BRCA1 gene product (BRCA1). The BRCA1
 CC protein has been characterized as a tumour suppressor protein.
 CC Alterations in the amino acid sequence of BRCA1 causes breast and ovarian
 CC cancers by removing the controls on cell growth and proliferation.
 CC Research has shown that different regions on the BRCA1 molecule have
 CC different effects on cell growth and tumour suppression (e.g. full length
 CC truncated BRCA1 has no effect on breast cancer cell growth but will
 CC inhibit ovarian cancer cell growth). It has been suggested that different
 CC host cell factors (e.g. proteins) interact with different regions of the
 CC BRCA1 to control its function. The identification of these proteins (e.g.
 CC BRCA1MP) will facilitate the development of novel diagnostic methods and
 CC new therapeutics for identifying and treating cancers caused by changes
 CC in the expression or activity of BRCA1
 XX
 SQ Sequence 386 AA;
 Query Match 3.6%; Score 19; DB 2; Length 386;
 Best Local Similarity 100.0%; Pred. No. 2.5e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 163 LNTLSTPEKLAALCKKYA 181
 DB 135 LNTLSTPEKLAALCKKYA 153
 RESULT 16
 ID ABO07222 standard; protein; 395 AA.
 AC ABO07222;
 XX
 XX 13-AUG-2003 (first entry)
 DT
 XX Human p53 modifying protein, SEQ ID 182.
 DE
 XX Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;
 KW antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
 KW lung cancer; ovarian cancer; angiogenesis; cell cycle;
 KW apoptotic disorder; cell proliferation disorder.
 XX
 XX Homo sapiens.
 OS
 XX WO200299122-A1.
 PN

DE Human ovarian antigen HLTJA50, SEQ ID NO:3413.
XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive.
XX Homo sapiens.
OS
XX
FN WO200200677-A1.
XX
XX 03-JAN-2002.
XX
XX 07-JUN-2001; 2001WO-US018559.
XX
XX 07-JUN-2000; 2000US-0209467P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CE, Rosen CA;
PI
XX
XX WPI; 2002-147878/19.
DR N-PSDB; ABQ55358.
XX
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
PT cancer), immune disorders, cardiovascular disorders and neurological
PT diseases.
XX
XX Claim 11; SEQ ID NO 3413; 2922pp; English.
PS
XX
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 85 AA;
3.2%; Score 17; DB 5; Length 85;
Query Match
Best Local Similarity 100.0%; Pred. No. 5.9e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OS

QY 397 QEMEKMTKKIKKLEKET 413
Db 52 QEMERMTKKIKKLEKET 68
|||||
RESULT 19
AAV17863
ID AAV17863 standard; protein; 676 AA.
AC AAV17863;
XX
DT 16-AUG-1999 (first entry)
XX
XX Neurite extending activity protein.
XX
XX Neurite extending activity; anti-dementia; memory; brain function;
KW dementia.
XX
XX Mus SP.
XX
FN JP11147897-A.
XX
PD 02-JUN-1999.
XX
XX 13-NOV-1997; 97JP-00331242.
PF
XX 13-NOV-1997; 97JP-00331242.
PR
XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
XX
XX WPI; 1999-379889/32.
DR N-PSDB; AAX80156.
XX
XX New protein with neurite extending activity - useful for treating
PT dementia.
PT
XX
XX Claim 1; Page 5-8; 9pp; Japanese.
PS
XX
XX The present sequence represents a protein (I) which has neurite extending
CC activity. (I) may be administered to patients to prevent dementia or to
CC improve memory and brain function. (I) is also used as a neurite-
CC extending agent
XX
XX Sequence 676 AA;
3.2%; Score 17; DB 2; Length 676;
Query Match
Best Local Similarity 100.0%; Pred. No. 3.4e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OS

QY 220 LARSKLESCLRELQREN 236
Db 207 LARSKLESCLRELQREN 223
|||||
RESULT 20
ABR43284
ID ABR43284 standard; protein; 684 AA.
AC ABR43284;
XX
XX 09-JUL-2003 (first entry)
DT
XX
XX Human neurotransmission-associated protein NTRAN-14 SEQ ID NO:14.
XX
XX Human; neurotransmission-associated protein; NTRAN; cytostatic; anti-HIV;
KW antiarteriosclerotic; anticonvulsant; nootropic; neuroprotective; stroke;
KW cerebrotective; antiallergic; antiinflammatory; thyromimetic; cancer;
KW antidiabetic; gene therapy; cell proliferative disorder; atherosclerosis;
KW neurological disorder; epilepsy; Huntington's disease; immune disorder;
KW inflammatory disorder; AIDS; allergy; developmental disorder; diabetes;
KW hypothyroidism; Cushing's syndrome; endocrine disorder; infection.
XX
XX Homo sapiens.
OS

```

XX WO2003025129-A2.
XX
XX 27-MAR-2003.
XX
XX 12-SEP-2002; 2002WO-US029219.
XX
XX 14-SEP-2001; 2001US-0322180P.
XX
XX 28-SEP-2001; 2001US-0326096P.
XX
XX 04-OCT-2001; 2001US-0327446P.
XX
XX 26-OCT-2001; 2001US-0345837P.
XX
XX 02-NOV-2001; 2001US-0343903P.
XX
XX 27-NOV-2001; 2001US-0334020P.
XX
XX 07-DEC-2001; 2001US-0340226P.
XX
XX 04-JAN-2002; 2002US-0345008P.
XX
XX 18-MAR-2002; 2002US-0365645P.
XX
XX 10-MAY-2002; 2002US-0379887P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Honchell CD, Warren BA, Borowsky ML, Griffin JA, Li JX, Lee SY;
XX Yue H, Forsythe IJ, Marquis JP, Gietzen KJ, Baughn MR, Tran UK;
XX Lehr-Mason PM, Tang YT, Ramkumar J, Emerling BM, Lee EA, Elliott VS;
XX Hafalia AJA, Duggan BM, Chawla NK, Kable AE, Chang H, Khare R;
XX Becha SD, Jin P, Lee S;
XX
XX WPI; 2003-363137/34.
XX
XX N-PSDB; ACC68992.
XX
XX New human neurotransmission associated proteins (NTRAN), useful for
XX diagnosing, treating and preventing diseases or conditions associated
XX with the aberrant NTRAN expression e.g. cancer, AIDS, diabetes, epilepsy,
XX or infections.
XX
XX Claim 1; Page 201-203; 240pp; English.
XX
XX ACC68992 to ACC69003 encode the human neurotransmission-associated
XX proteins given in ABR43271 to ABR43295, designated NTRAN-1 to NTRAN-25
XX (1). (i) have cytostatic, antiarteriosclerotic, anticonvulsant,
XX neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic,
XX antiinflammatory, thymometric and antidiabetic activities, and can be
XX used in gene therapy. The NTRAN polypeptides and polynucleotides are
XX useful in diagnosing, treating and preventing diseases or conditions
XX associated with the decreased expression or overexpression of NTRAN, such
XX as cell proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
XX epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
XX allergies), developmental (e.g. hypothyroidism, Cushing's syndrome) or
XX endocrine (e.g. diabetes) disorders, or infections. They are also useful
XX in assessing the effects of exogenous compounds on the expression of
XX nucleic acid and amino acid sequences of NTRAN. The NTRANs or their
XX fragments are useful in screening compounds for effectiveness as agonist
XX or antagonist of the polypeptides, or in altering the expression of the
XX target polynucleotide and compounds that specifically bind to or modulate
XX the activity of the polypeptide
XX
XX Sequence 684 AA;
XX
XX Query Match 3.2%; Score 17; DB 6; Length 684;
XX Best Local Similarity 100.0%; Pred. No. 3.4e-07;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 220 LARSKLESCLRELQHN 236
XX |||||
XX DB 209 LARSKLESCLRELQHN 225
XX |||||
XX
XX RESULT 21
XX AA008301
XX ID AA008301 standard; protein; 475 AA.
XX
XX AC AA008301;
XX
XX DT 06-NOV-2001 (first entry)
XX

```

```

XX Human polypeptide SEQ ID NO 22193.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US004927.
XX
XX 28-FEB-2000; 2000US-00515136.
XX
XX 18-MAY-2000; 2000US-00577409.
XX
XX (HYSB-) HYSSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX N-PSDB; AAI88232.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX and treating e.g. leukemia, inflammation and immune disorders.
XX
XX Claim 20; SEQ ID NO 22193; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ffp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 475 AA;
XX
XX Query Match 2.3%; Score 12; DB 4; Length 475;
XX Best Local Similarity 100.0%; Pred. No. 0.017;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 245 QRAREEEKKRKE 256
XX |||||
XX DB 33 QRAREEEKKRKE 44
XX
XX RESULT 22
XX ADC3216
XX ID ADC3216 standard; protein; 475 AA.
XX
XX AC ADC3216;
XX
XX 18-DEC-2003 (first entry)
XX
XX Human novel contig-encoded polypeptide sequence, SEQ ID NO:3298.
XX
XX Human; diagnostic; drug screening; forensics; gene mapping;
XX biodiversity assessment; Parkinson's disease; Alzheimer's disease;
XX neurodegenerative diseases; anaemia; platelet disorder; wound; Burns;
XX ulcers; osteoporosis; autoimmune disease; cancer;
XX molecular weight marker; food supplement; antiparkinsonian; neurotropic;
XX neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnery;
XX antilulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;

```

gene therapy; chromosome 6.
Homo sapiens.
WO2003029271-A2.
10-APR-2003.
24-SEP-2002; 2002WO-US030474.
24-SEP-2001; 2001US-0324631P.
(HYSE-) HYSEQ INC.
Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
Zhou P, Chosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
Haley-Vicente D, Drmanac RT;
WPI; 2003-371981/35.
N-PSDB; ADC32449.
New polynucleotide and polypeptide useful for diagnosing, preventing or
treating conditions such as neurodegenerative diseases, anemias, platelet
disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
cancer.
Example 2; SEQ ID NO 3298; 1185pp; English.
The invention relates to 971 novel human cDNA sequences (ADC29919-
ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
invention also relates to nucleic acid sequences over 99% identical with
the novel human cDNAs. The invention additionally encompasses expression
vectors and host cells comprising a nucleic acid of the invention; the
recombinant production of a polypeptide of the invention; an antibody
against a polypeptide of the invention; a method of detecting
polynucleotides or polypeptides of the invention; and methods of
identifying a compound which binds to a polypeptide of the invention. The
invention further discloses methods of preventing, treating or
ameliorating a medical condition; kits comprising polynucleotide probes
and/or monoclonal antibodies for carrying out the methods of the
invention; methods for the identification of compounds that modulate the
expression or activity of the polynucleotide and/or polypeptide; and 767
contig sequences corresponding to the cDNA sequences of the invention
(ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
-ADC33394). The nucleic acids and polypeptides of the invention are
useful in diagnostics, drug screening, forensics, gene mapping, in the
identification of mutations responsible for genetic disorders or other
traits, for assessing biodiversity, and in producing many other types of
data and products dependent on DNA and amino acid sequences. They are
also used for treating diseases such as Parkinson's disease, Alzheimer's
disease and other neurodegenerative diseases, anaemia, platelet
disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
cancer. The nucleic acids may also be used as hybridisation probes or
primers, and in the recombinant production of a protein. The polypeptides
are also useful in generating antibodies, as molecular weight markers,
and as food supplements. The present sequence represents a human contig-
encoded polypeptide sequence used in an example of the invention. Note:
The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
Sequence 475 AA;
Query Match
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 245 QPAREEERKKE 256
DB 33 QPAREEERKKE 44
RESULT 23

AAU22308
ID AAU22308 standard; protein; 99 AA.
XX
AC AAU22308;
XX
DT 18-DEC-2001 (first entry)
XX
XX Human cardiovascular system antigen polypeptide SEQ ID NO 1082.
XX
XX Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
XX chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
XX antirheumatic; antiproliferative; cytostatic; cardant; neuroprotective;
XX cerebroprotective; nontropic; antibacterial; virucide; fungicide; cancer;
XX ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
XX hyperproliferative disorder; breast; liver; cardiovascular disorder;
XX cerebrovascular disorder; nervous system disorder; bacterial infection;
XX fungal infection; viral infection; ocular disorder; endocrine disorder;
XX gastrointestinal disorder; renal disorder; respiratory disorder;
XX wound healing; skin aging; organ transplantation; tissue regeneration;
XX anti-infertility.
XX
XX Homo sapiens.
XX
XX WO200155321-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001340.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184564P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214866P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 11-JUL-2000; 2000US-0217487P.
XX 14-JUL-2000; 2000US-0217496P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226861P.
XX 22-AUG-2000; 2000US-0226868P.
XX 23-AUG-2000; 2000US-0227182P.
XX 30-AUG-2000; 2000US-0227009P.
XX 01-SEP-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.

```
PR 08-SEP-2000; 2000US-0231342P.
PR 08-SEP-2000; 2000US-0231343P.
PR 08-SEP-2000; 2000US-0231344P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231368P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234978P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239315P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241222P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.

PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 08-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-451930/48.
XX N-PSDB; AAS35582.
XX
XX New cardiovascular system related polynucleotides and polypeptides,
PT useful for diagnosing, treating and/or preventing disorders of the
PT cardiovascular system.
XX
XX Claim 11; SEQ ID NO 1082; 674pp; English.
XX
XX Sequences AAU21852-AAU22466 represent the cardiovascular system antigen
CC polypeptides of the invention. Cardiovascular system antigens and their
CC associated polynucleotides are useful in the diagnosis, treatment and
CC prevention of various types of disorders in e.g. humans, mice, rabbits,
CC goats, horses, cats, dogs, chickens or sheep. A pathological condition
CC can be determined by detecting the presence or absence of a mutation in a
CC cardiovascular system antigen polynucleotide. The treatable disorders
CC include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma and pleurisy. The polypeptides can
CC also be used to aid wound healing, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, to regenerate tissues and in
CC chemotaxis. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

Query Match 1.7%; Score 9; DB 4; Length 99;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GEKPSKGD 123
DB 49 GEKPSKGD 57
|||||
|

RESULT 24
ADE46276
ID ADE46276 standard; protein; 99 AA.
XX
XX ADE46276;
XX
XX 29-JAN-2004 (first entry)
XX
```


PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764869.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Ruben SM, Barash SC;
XX WPI; 2003-743766/70.
XX N-PSDB; ADE45661.
XX New cardiovascular system related polynucleotides and polypeptides,
PT useful for preventing, treating, or ameliorating a medical condition,
PT such as cancer of cardiovascular tissues and cancer metastases.
XX Claim 11; SEQ ID NO 1082; 262pp; English.
XX The invention relates to human cardiovascular system related polypeptides
CC and the polynucleotides encoding them. The polypeptides, polynucleotides
CC and antibodies to the polypeptides are useful for diagnosing a
CC pathological condition or a susceptibility to a pathological condition,
CC for preventing, treating, or ameliorating a medical condition, such as
CC cancer of cardiovascular system tissues, proliferative disorders, foetal
CC cancer of cardiovascular system tissues, haematopoietic disorders, diseases of
CC the immune system, AIDS, autoimmune diseases (e.g. rheumatoid
CC arthritis), inflammation, allergies, neurological disorders (e.g.,
CC Alzheimer's disease, Parkinson's disease), cognitive disorders,
CC schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis,
CC diabetes, atherosclerosis, cardiovascular disorders, angiogenic
CC disorders, kidney disorders, gastrointestinal disorders, pregnancy-
CC related disorders, endocrine disorders and infections. The nucleic acids
CC are also useful for chromosome identification, radiation hybrid mapping
CC or long-range restriction mapping. The polypeptides and polynucleotides
CC may also be used as food additives or preservatives to increase or
CC decrease storage capabilities, fat content or other nutritional
CC components. This sequence represents a human cardiovascular system
CC related polypeptide of the invention.
XX Sequence 99 AA;
SQ
Query Match 1.7%; Score 9; DB 7; Length 99;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 115 GEKEPSKGD 123
DB 49 GEKEPSKGD 57
RESULT 25
ABBS59454
ID ABB59454 standard; protein; 554 AA.
XX ABB59454;
AC ABB59454;
XX 26-MAR-2002 (first entry)
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 5154.
D3 Drosophila melanogaster.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
OS Drosophila melanogaster.
XX Drosophila melanogaster.
PI WO200171042-A2.

XX 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEXE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI RPI; 2001-656860/75.
DR N-PSDB; ABL03557.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Disclosure; SEQ ID NO 5154; 21pp + Sequence Listing; English.
PS The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 554 AA;
SQ
Query Match 1.7%; Score 9; DB 4; Length 554;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 248 REEEKRKE 256
DB 488 REEEKRKE 496
RESULT 26
ABU44191
ID ABU44191 standard; protein; 637 AA.
XX ABU44191;
AC ABU44191;
XX 19-JUN-2003 (first entry)
DT 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #29718.
DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Streptococcus mutans.
OS Streptococcus mutans.
XX WO200277183-A2.
PN WO200277183-A2.
XX 03-OCT-2002.
PD 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
PF 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
PA Wang L, Zamudio C, Malone C, Haseibeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
PI

```

XX WPI; 2003-029926/02.
DR N-PSDB; ACA48061.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 72115; 1766bp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 637 AA;
XX
XX Query Match 1.7%; Score 9; DB 6; Length 637;
XX Best Local Similarity 100.0%; Pred. No. 17;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 57 BELSRQLED 65
XX |||||
XX 629 BELSRQLED 637
XX
XX RESULT 27
XX AAR89275
XX ID AAR89275 standard; protein; 729 AA.
XX
XX AC AAR89275;
XX
XX 25-JUL-1996 (first entry)
XX
XX Yeast coagulation protein FLO8.
XX
XX Saccharomyces cerevisiae; coagulation; FLO8; mutant; wild type; wine;
XX beer; yeast; flavour; alcoholic beverage.
XX
XX Saccharomyces cerevisiae.
XX
XX JP08000270-A.
XX
XX 09-JAN-1996.
XX
XX 23-JUN-1994; 94JP-00141520.
XX
XX 23-JUN-1994; 94JP-00141520.
XX (KIRI ) KIRIN BREWERY KK.
XX
XX WPI; 1996-091654/10.
XX N-PSDB; AAR99257.
XX
XX Yeast coagulating gene FLO8 and prepn. of yeast with increased
PT coagulation - also prepn. of yeast with decreased coagulation, both
PT useful in wine making and brewing industries.
XX
XX Claim 1; Page 13-16; 18pp; Japanese.
XX
XX This is the amino acid sequence of the Saccharomyces cerevisiae
CC coagulation protein FLO8. The corresp. gene was isolated by screening a
CC non-coagulating yeast strain (YPH500) transformed with a library
CC generated from DNA from a coagulating yeast strain (ARCC-60715). The full
CC length gene can be used to generate mutants having reduced or increased
CC coagulation activity. The strain carrying the wild type or mutant FLO8
CC genes can be used in wine and beer making, where the coagulating activity
CC of the yeast can affect the flavour of the alcoholic beverage
XX
XX Sequence 729 AA;
XX
XX Query Match 1.7%; Score 9; DB 2; Length 729;
XX Best Local Similarity 100.0%; Pred. No. 19;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 164 NTLSTPEEK 172
XX |||||
XX 601 NTLSTPEEK 609
XX
XX RESULT 28
XX AAY12665
XX ID AAY12665 standard; protein; 65 AA.
XX
XX AC AAY12665;
XX
XX 22-JUN-1999 (first entry)
XX
XX Human 5' EST secreted protein SEQ ID NO: 330 from WO 9906553.
XX
XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX forensic; gene therapy; chromosome mapping; signal peptide;
XX upstream regulatory sequence; cytokine activity; cell proliferation;
XX differentiation; haematopoiesis regulation; tissue growth regulation;
XX reproductiv hormone regulation; chemotactic; chemokinetic; haemostatic;
XX thrombolytic; antiinflammatory; tumour inhibition; antitumour.
XX
XX Homo sapiens.
XX
XX OS
XX WO9906553-A2.
XX
XX 11-FEB-1999.
XX
XX 31-JUL-1998; 98WO-IB001237.
XX
XX 01-AUG-1997; 97US-00905051.
XX (CBST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Lacroix B;
XX
XX WPI; 1999-153783/13.
XX N-PSDB; AAX41523.
XX
XX New nucleic acids encoding human secreted proteins - obtained from cDNA
XX libraries derived from umbilical cord, lymph ganglia, lymphocytes and
XX placental tissue.
XX
XX Claim 34; Page 408; 411pp; English.

```


XX AAX41379 to AAX41526 represent 5' expressed sequence tags (ESTs) for
 CC human secreted proteins, and encode the proteins given in AAY12521 to
 CC AAY12668, respectively. The proteins given represent the signal peptide
 CC and an N-terminal fragment of a secreted protein. The nucleic acid
 CC sequences can be used for producing secreted human gene products. They
 CC can also be used to develop products for diagnosis and therapy. The
 CC proteins obtained may have cytokine activity, cell
 CC proliferation/differentiation activity, haematopoiesis regulating
 CC activity, tissue growth regulating activity, reproductive hormone
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
 CC thrombolytic activity, receptor/ligand activity, antiinflammatory
 CC activity, tumour inhibition activity or other activities. The products
 CC can be used in forensic, gene therapy and chromosome mapping procedures.
 CC The sequences can also be used for obtaining corresponding promoter
 CC sequences. The nucleic acids encoding the signal peptide can be used for
 CC directing extracellular secretion of a polypeptide or the insertion of a
 CC polypeptide into a membrane, or importing a polypeptide into a cell
 CC
 XX Sequence 65 AA;
 SQ

Query Match 1.5%; Score 8; DB 2; Length 65;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 492 PGQAQPS 499
 DB 10 PGQAQPS 17
 |||||

RESULT 29
 AAY01195
 ID AAY01195 standard; protein; 156 AA.
 XX
 AC AAY01195;
 XX
 DT 18-MAY-1999 (first entry)
 XX
 DE Polypeptide fragment encoded by gene 15.
 XX
 KW Human; secreted protein; gene therapy; protein therapy; tissue; cancer;
 KW tumour; neurodegenerative disorder; leukaemia; autoimmune disease; AIDS;
 KW developmental abnormality; foetal deficiency; Alzheimer's disease;
 KW cognitive disorder; schizophrenia; immunological disorder; mood disorder;
 KW immune deficiency disease; respiratory disorder; arthritis; skeletal;
 KW haematopoietic disorder; neural; osteoporosis; metabolic disorders;
 KW cardiovascular; endocrine; gastrointestinal; asthma; diagnosis.
 XX
 CS Homo sapiens.
 XX
 FN WO9901020-A2.
 XX
 PD 14-JAN-1999.
 XX
 PF 30-JUN-1998; 98WO-US013608.
 XX
 PR 01-JUL-1997; 97US-0051381P.
 PR 01-JUL-1997; 97US-0051480P.
 PR 12-SEP-1997; 97US-0058398P.
 PR 12-SEP-1997; 97US-0058663P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Carter KC, Feng P, Rosen CA, Ruben SM, Endress GA;
 PI
 DR WPI; 1999-105683/09.
 DR N-PSDB; AAX22125.
 DR
 XX New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, immune deficiency diseases or blood
 PT disorders.
 PT

PS
 XX Disclosure; Page 26; 179pp; English.
 CC
 CC The invention relates to nucleic acid sequences (AAX22111 to AAX22134)
 CC encoding human secreted proteins (AAY01135 to AAY01158). The secreted
 CC protein gene sequences are deposited with the ATCC under deposit number
 CC ATCC 29918. Host cells comprising recombinant vectors containing the
 CC nucleic acid sequences are used for the recombinant production of the
 CC secreted proteins. The polynucleotide and amino acid sequences are useful
 CC for are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Pathological conditions can
 CC be also diagnosed by determining the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new
 CC polynucleotides. Specific uses are described for each of the
 CC polynucleotides, based on which tissues they are most highly expressed
 CC in, and include developing products for the diagnosis or treatment of
 CC cancer, tumours, developmental abnormalities and foetal deficiencies,
 CC autoimmune diseases, lymphomas, Alzheimer's and cognitive disorders,
 CC schizophrenia, immunological disorders, immune deficiency diseases
 CC (AIDS), mood disorders, respiratory disorders, arthritis, asthma,
 CC haematopoietic disorders, neural disorders, skeletal disorders,
 CC osteoporosis, metabolic disorders, cardiovascular disorders, endocrine
 CC disorders or gastrointestinal disorders. The polypeptides are also useful
 CC for identifying their binding partners. The present sequence represents a
 CC polypeptide fragment encoded by a gene of the invention (see descriptor
 CC line for gene number)
 XX
 SQ Sequence 156 AA;
 Query Match 1.5%; Score 8; DB 2; Length 156;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 468 RVQDLSAG 475
 DB 87 RVQDLSAG 94
 |||||

RESULT 30
 AAG08727
 ID AAG08727 standard; protein; 170 AA.
 XX
 AC AAG08727;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 6379.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN BP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-00301439.
 XX
 PR 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 23-MAR-1999; 99US-0123548P.
 PR 25-MAR-1999; 99US-0125789P.
 PR 29-MAR-1999; 99US-0126264P.
 PR 01-APR-1999; 99US-0126783P.
 PR 06-APR-1999; 99US-0127462P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130445P.
 PR 23-APR-1999; 99US-0130510P.
 PR 23-APR-1999; 99US-0130891P.

PR 23-JUL-1999; 99US-0145145P.
 PR 23-JUL-1999; 99US-0145218P.
 PR 23-JUL-1999; 99US-0145224P.
 PR 26-JUL-1999; 99US-0145276P.
 PR 27-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0145951P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146388P.
 PR 02-AUG-1999; 99US-0146389P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 04-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147192P.
 PR 05-AUG-1999; 99US-0147260P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 06-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147935P.
 PR 10-AUG-1999; 99US-0148171P.
 PR 11-AUG-1999; 99US-0148319P.
 PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149426P.
 PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149929P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151080P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 10-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154033P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0155659P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-015659P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-015753P.
 PR 06-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 18-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.

PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 23-OCT-1999; 99US-0161404P.
 PR 23-OCT-1999; 99US-0161405P.
 PR 23-OCT-1999; 99US-0161406P.
 PR 23-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161922P.
 PR 28-OCT-1999; 99US-0161933P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 1.5%; Score 8; DB 3; Length 174;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 AAPAAVEAE 28
 |||||
 DB 34 AAPAAVEAE 41

RESULT 32

ABG19843
 ID ABG19843 standard; protein; 194 AA.

XX AC ABG19843;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #19834.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WC200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB; AAS84030.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits and to assess
 biodiversity.

XX PS Claim 20; SEQ ID NO 50202; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 sequences. (I) is useful as hybridisation probes, polymerase chain
 reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 and in recombinant production of (II). The polynucleotides are also used
 in diagnostics as expressed sequence tags for identifying expressed
 genes. (I) is useful in gene therapy techniques to restore normal
 activity of (II) or to treat disease states involving (II). (II) is
 useful for generating antibodies against it, detecting or quantitating a
 polypeptide in tissue, as molecular weight markers and as a food
 supplement. (II) and its binding partners are useful in medical imaging
 of sites expressing (II). (I) and (II) are useful for treating disorders
 involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 194 AA;

Query Match 1.5%; Score 8; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 476 GQGLTDS 483
Db 136 GQGLTDS 143
|||||

RESULT 33

AAAG20570
ID AAAG20570 standard; protein; 224 AA.

AC AAAG20570;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 22814.

XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-03101439.

XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126284P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132407P.

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 07-MAY-1999; 99US-0132487P.

PR 11-MAY-1999; 99US-0132863P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 18-MAY-1999; 99US-0134370P.

PR 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.

PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.

```
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148313P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151068P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153707P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157553P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158363P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161408P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.

PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 1.5%; Score 8; DB 3; Length 224;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 AAPAVEAE 28
Db 84 AAPAVEAE 91

RESULT 34
ABU05665
ID ABU05665 standard; protein; 229 AA.
XX AC ABU05665;
XX DT 08-APR-2003 (first entry)
XX M. tuberculosis and M. leprae marker protein #316.
DE XX Mycobacterioses; survival; virulence; protective antigen; vaccine;
KW Mycobacterial disease; tuberculosis; leprosy.
XX OS Mycobacterium tuberculosis.
XX OS Mycobacterium leprae.
XX FN WO200274903-A2.
XX PD 26-SEP-2002.
XX PP 22-FEB-2002; 2002WO-IB001973.
XX PR 22-FEB-2001; 2001US-0270123P.
XX PA (INSP ) INST PASTEUR.
XX PI Cole S;
XX DR WPI; 2002-759985/82.
XX PT Identifying and selecting genes for survival or virulence of mycobacteria
PT by a comparative genomic analysis of the sequences of Mycobacterium
PT tuberculosis and M. leprae.
XX PS Claim 17; Page 500-501; 874pp; English.
XX CC This invention relates to a novel method for identifying essential genes
CC for survival or virulence of mycobacteria species. The method comprises
CC aligning the genomic sequence of a first mycobacterium species on a
CC genomic sequence of a second mycobacterium species and selecting a
CC polynucleotide sequence that is highly conserved in both genomes with no
CC counterparts in other bacterial genomic sequences and that corresponds to
CC an essential gene for the survival or virulence of mycobacterium species.
CC The method of the invention is useful for detecting M. tuberculosis or M.
CC leprae infection. The method reduces the number of potential new targets
CC and protective antigens for new drugs and vaccine compositions to treat
CC and prevent mycobacterial diseases, particularly tuberculosis and
CC leprosy. The present sequence represents a marker protein from
CC Mycobacterium tuberculosis and Mycobacterium leprae identified using the
CC method of the invention
XX SQ Sequence 229 AA;

Query Match 1.5%; Score 8; DB 5; Length 229;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 PSQAPAV 25
|||||
```

```
Db      200 PQAAPAV 207

RESULT 35
ABG77243
ID ABG77243 standard; protein; 360 AA.
XX
AC ABG77243;
XX
DT 05-NOV-2002 (first entry)
XX
DE Selected Interacting Domain (SID) polypeptide #54.
XX
KW Yeast; selected interacting domain; SID; antifungal; cancer; cytostatic;
KW neuroprotective; Candida infection; gene therapy;
KW neurodegenerative disease.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200259255-A2.
XX
PD 01-AUG-2002.
XX
PF 25-JAN-2002; 2002WO-EP001350.
XX
PR 26-JAN-2001; 2001US-0264577P.
XX
PA (HYBR-) HYBRIGENICS.
XX
PI Legrain P;
XX
DR WPI; 2002-619165/66.
DR N-PSDB; ABS62857.
XX
XX New complex between two interacting bait and prey Saccharomyces
PT cerevisiae polypeptides, useful for preventing or treating Candida
PT infection, cancer or neurodegenerative diseases in a mammal.
XX
PS Claim 6; Page 77; 196pp; English.
XX
CC The invention relates to a complex between two interacting Saccharomyces
CC cerevisiae polypeptides, comprising two Selected Interacting Domain (SID)
CC polypeptides as bait and prey proteins. A pharmaceutical composition
CC comprising the complex is useful for preventing or treating Candida
CC infection, cancer and neurodegenerative diseases in a human or animal,
CC preferably in a mammal. This sequence represents a SID polypeptide of the
CC invention
XX
XX Sequence 360 AA;
XX
Query Match 1.5%; Score 8; DB 5; Length 360;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEKKKE 256
DB 240 EEEKKKE 247
|||||
RESULT 37
ABJ11057
ID ABJ11057 standard; protein; 364 AA.
XX
AC ABJ11057;
XX
DT 10-DEC-2002 (first entry)
XX
DE Yeast selected interacting domain protein SEQ ID NO: 248.
XX
KW Yeast; protein-protein interaction; Selected Interacting Domain;
KW SID (RTM); secretion yield; cancer; neurodegenerative disease; fungicide;
KW cytostatic; neuroprotective.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200266504-A2.
XX
PD 29-AUG-2002.
XX
PF 14-FEB-2002; 2002WO-EP002299.
XX
PR 16-FEB-2001; 2001US-0269266P.
XX
PA (HYBR-) HYBRIGENICS.
XX
PI Legrain P;
XX

Saccharomyces cerevisiae.
WO200266504-A2.
29-AUG-2002.
14-FEB-2002; 2002WO-EP002299.
16-FEB-2001; 2001US-0269266P.
(HYBR-) HYBRIGENICS.
Legrain P;
WPI; 2002-674913/72.
N-PSDB; ABT11431.
New protein-protein complexes of Saccharomyces cerevisiae, useful in drug
screening or development, for developing yeast strains with better
secretion yield of protein, or in gene therapy (e.g. to treat Candida
infection or cancer).
Claim 6; Page 164; 357pp; English.
The present invention relates to complexes between Saccharomyces
cerevisiae Selected Interacting Domain (SID (RTM)) proteins and coding
sequences. The protein complexes of S. cerevisiae are useful in drug
development, in screening drugs or agents that modulate the interaction
of proteins, for developing yeast strains with better secretion yield of
protein, and in gene therapy. The protein complexes, polypeptides and
polynucleotides are useful for preventing or treating Candida infection,
cancer or neurodegenerative diseases in humans or animals. The present
sequence is a protein of the invention
XX
XX Sequence 360 AA;
XX
Query Match 1.5%; Score 8; DB 5; Length 360;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEKKKE 256
DB 240 EEEKKKE 247
|||||
RESULT 37
ABJ11057
ID ABJ11057 standard; protein; 364 AA.
XX
AC ABJ11057;
XX
DT 10-DEC-2002 (first entry)
XX
DE Yeast selected interacting domain protein SEQ ID NO: 248.
XX
KW Yeast; protein-protein interaction; Selected Interacting Domain;
KW SID (RTM); secretion yield; cancer; neurodegenerative disease; fungicide;
KW cytostatic; neuroprotective.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200266504-A2.
XX
PD 29-AUG-2002.
XX
PF 14-FEB-2002; 2002WO-EP002299.
XX
PR 16-FEB-2001; 2001US-0269266P.
XX
PA (HYBR-) HYBRIGENICS.
XX
PI Legrain P;
XX
```


CC The invention relates to 971 novel human cDNA sequences (ADC29919-
 CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
 CC invention also relates to nucleic acid sequences over 99% identical with
 CC the novel human cDNAs. The invention additionally encompasses expression
 CC vectors and host cells comprising a nucleic acid of the invention; the
 CC recombinant production of a polypeptide of the invention; an antibody
 CC against a polypeptide of the invention; a method of detecting
 CC polynucleotides or polypeptides of the invention; and methods of
 CC identifying a compound which binds to a polypeptide of the invention. The
 CC invention further discloses methods of preventing, treating or
 CC ameliorating a medical condition; kits comprising polynucleotide probes
 CC and/or monoclonal antibodies for carrying out the methods of the
 CC invention; methods for the identification of compounds that modulate the
 CC expression or activity of the polynucleotide and/or polypeptide; and 767
 CC contig sequences corresponding to the cDNA sequences of the invention
 CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
 CC -ADC33394). The nucleic acids and polypeptides of the invention are
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the
 CC identification of mutations responsible for genetic disorders or other
 CC traits, for assessing biodiversity, and in producing many other types of
 CC data and products dependent on DNA and amino acid sequences. They are
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, anaemia, platelet
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 CC cancer. The nucleic acids may also be used as hybridisation probes or
 CC primers, and in the recombinant production of a protein. The polypeptides
 CC are also useful in generating antibodies, as molecular weight markers,
 CC and as food supplements. The present sequence represents a specifically
 CC claimed human polypeptide sequence of the invention. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 416 AA;

Query Match 1.5%; Score 8; DB 7; Length 416;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 299 QVELREEH 306

Db 20 QVELREEH 27

RESULT 40

AAE79356

ID AAE79356 standard; protein; 484 AA.

AC AAE79356;

DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:228.

XX Corynebacterium glutamicum; carbon metabolism and energy production;
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.

XX Corynebacterium glutamicum.

XX WO200100844-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WC-IB000943.

XX 25-JUN-1999; 99US-0141031P.

XX 08-JUL-1999; 99DE-01031412.

XX 08-JUL-1999; 99DE-01031413.

PR 08-JUL-1999; 99DE-01031419.
 PR 08-JUL-1999; 99DE-01031420.
 PR 08-JUL-1999; 99DE-01031424.
 PR 08-JUL-1999; 99DE-01031428.
 PR 08-JUL-1999; 99DE-01031431.
 PR 08-JUL-1999; 99DE-01031433.
 PR 08-JUL-1999; 99DE-01031434.
 PR 08-JUL-1999; 99DE-01031510.
 PR 08-JUL-1999; 99DE-01031562.
 PR 08-JUL-1999; 99DE-01031634.
 PR 09-JUL-1999; 99DE-01032180.
 PR 09-JUL-1999; 99DE-01032227.
 PR 09-JUL-1999; 99DE-01032230.
 PR 09-JUL-1999; 99US-0143208P.
 PR 14-JUL-1999; 99DE-01032924.
 PR 14-JUL-1999; 99DE-01032973.
 PR 14-JUL-1999; 99DE-01033005.
 PR 27-AUG-1999; 99DE-01040765.
 PR 31-AUG-1999; 99US-0151572P.
 PR 03-SEP-1999; 99DE-01042076.
 PR 03-SEP-1999; 99DE-01042086.
 PR 03-SEP-1999; 99DE-01042087.
 PR 03-SEP-1999; 99DE-01042088.
 PR 03-SEP-1999; 99DE-01042095.
 PR 03-SEP-1999; 99DE-01042123.
 PR 03-SEP-1999; 99DE-01042125.

XX (BADI) BASF AG.

PA Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

PI WPI: 2001-061975/07.

DR N-PSDB; AAF71473.

XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT metabolism of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes.

XX Claim 20; Page 464-466; 1246pp; English.

XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in AAB79243
 CC to AAB 79633 which are involved in carbon metabolism and energy
 CC production. The C. glutamicum SMP gene can be used in vectors (II) for
 CC expression in host cells and production or modulation of production of
 CC fine chemicals, such as, an organic acid, a proteinogenic or
 CC nonproteinogenic amino acid (preferred), a purine or pyrimidine base, a
 CC nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty acid,
 CC a diol, a carbohydrate, an aromatic compound, a vitamin, a cofactor, a
 CC polypeptide, or an enzyme. The presence of (I) or SMP proteins (III)
 CC encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to C.
 CC glutamicum, identify and localise C. glutamicum sequences of interest, in
 CC evolutionary studies, in determining SMP protein regions required for
 CC function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH)

XX Sequence 484 AA;

Query Match 1.5%; Score 8; DB 4; Length 484;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 AELLEERH 188

Db 80 AELLEERH 87

RESULT 41

AAG92676
 ID AAG92676 standard; protein; 484 AA.
 XX
 AC AAG92676;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 6430.
 XX
 DE C glutamicum protein fragment SEQ ID NO: 6430.
 XX
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EPI108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 XX 18-DEC-2000; 2000EP-00127688.
 XX
 XX 16-DEC-1999; 99JP-00377484.
 XX
 PR 07-APR-2000; 2000JP-00159162.
 PR
 PR 03-AUG-2000; 2000JP-00280988.
 PR
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 PI
 XX WPI; 2001-376931/40.
 DR
 DR N-PSDB; AAF67895.
 XX
 XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analyzing
 PT expression profile or pattern of a gene and identifying homologous gene.
 PT
 XX Claim 17; SEQ ID NO 6430; 246pp + Sequence Listing; English.
 ES
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium glutamicum glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium, measuring expression amount and analysing
 CC the expression profile or expression pattern of a gene derived from
 CC Corynebacterium, and identifying a homologue of a gene derived from
 CC Corynebacterium. Corynebacterium bacteria are useful for producing amino
 CC acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described in the
 CC exemplification of the invention. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from the European Patent Office
 XX
 XX Sequence 484 AA;
 SQ
 Query Match 1.5%; Score 8; DB 4; Length 484;
 Best Local Similarity 100.0%; Pred.No. 1.3e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 181 AELLLEHR 188
 DB 80 AELLLEHR 87
 |||||
 RESULT 42
 AAG41279
 ID AAG41279 standard; protein; 795 AA.
 XX
 AC AAG41279;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 51338.
 DE
 XX Protein identification; signal transduction pathway; metabolic pathway;
 KW

KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS
 PN EPI033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-00301439.
 PF
 XX 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0125785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128334P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 28-APR-1999; 99US-0130891P.
 PR 30-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 04-MAY-1999; 99US-0132407P.
 PR 05-MAY-1999; 99US-0132484P.
 PR 06-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 07-MAY-1999; 99US-0132487P.
 PR 11-MAY-1999; 99US-0132863P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136021P.
 PR 27-MAY-1999; 99US-0136392P.
 PR 28-MAY-1999; 99US-0136782P.
 PR 01-JUN-1999; 99US-0137222P.
 PR 03-JUN-1999; 99US-0137528P.
 PR 04-JUN-1999; 99US-0137502P.
 PR 07-JUN-1999; 99US-0137724P.
 PR 08-JUN-1999; 99US-0138094P.
 PR 10-JUN-1999; 99US-0138540P.
 PR 10-JUN-1999; 99US-0138847P.
 PR 14-JUN-1999; 99US-0139119P.
 PR 16-JUN-1999; 99US-0139452P.
 PR 16-JUN-1999; 99US-0139453P.
 PR 17-JUN-1999; 99US-0139492P.
 PR 18-JUN-1999; 99US-0139454P.
 PR 18-JUN-1999; 99US-0139455P.
 PR 18-JUN-1999; 99US-0139456P.
 PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139763P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 22-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 23-JUN-1999; 99US-0140354P.

```
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143824P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 09-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 26-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159323P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 22-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
PR 29-OCT-1999; 99US-0162143P.
PR 29-OCT-1999; 99US-0162144P.

Query Match 1.5%; Score 8; DB 3; Length 785;
Best Local Similarity 100.0%; Pred. No. 1.9e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY 127 BEIRQSD 134
Db 216 BEIRQSD 223
|||||
|||||

RESULT 43
AAG41278
ID AAG41278 standard; protein; 823 AA.
XX
AC AAG41278;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51337.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
```

PD 06-SEP-2000. 99US-0142390P.
XX 08-JUL-1999; 99US-0142803P.
PP 09-JUL-1999; 99US-0142920P.
XX 12-JUL-1999; 99US-0143977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0148684P.
PR 17-AUG-1999; 99US-0149368P.
PR 18-AUG-1999; 99US-0149175P.
PR 20-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.

PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0155596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0158295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 1.5%; Score 8; DB 3; Length 823;
Best Local Similarity 100.0%; Pred. No. 2e+02; 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 EIRQSD 134
Db 254 EIRQSD 261

RESULT 44
AAG41277
ID AAG41277 standard; protein; 836 AA.

XX AC AAG41277;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 51336.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.

XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

XX PR 05-MAR-1999; 99US-0123180P.

XX PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 28-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135628P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.

PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0144586P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0145386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 09-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 10-AUG-1999; 99US-0147935P.
PR 11-AUG-1999; 99US-0148171P.
PR 12-AUG-1999; 99US-0148319P.
PR 13-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0148684P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 25-AUG-1999; 99US-0149930P.
PR 26-AUG-1999; 99US-0150566P.
PR 27-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.

PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 1.5%; Score 8; DB 3; Length 836;
Best Local Similarity 100.0%; Pred. No. 2e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0

QY 127 REIROSDE 134
Db |||||
267 REIROSDE 274

RESULT 45

ADB95084
ID ADB95084 standard; protein; 952 AA.
XX
AC ADB95084;
XX
DT 04-DEC-2003 (first entry)
XX
DE A. thaliana protein 62837 #SEQ ID 82.
XX
KW Plant; herbicide; weed; crop field; growth; development.
XX OS Arabidopsis thaliana.
XX PN WC2003008440-A2.
XX
XX 30-JAN-2003.
XX
XX 16-JUL-2002; 2002WC-EP007929.
PR 16-JUL-2001; 2001US-0305806P.
PR 20-FEB-2002; 2002US-0358416P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
XX Levin JZ, Patton DA, McElver JA, Budziszewski GJ, Zhou Q, Aux GW;
PI Tossberg J, Wegrich Glover L, Ashby CS, Thomas CR, Madhaven E;
PI Lewis S, Dunn J, Cates E, Law MD;
XX
XX WPI; 2003-229557/22.
DR N-PSDB; ADB95083.
XX

.....

XX PA (CURA-) CURAGEN CORP.
XX PI Shimkets RA, Leach M;
XX PT WPI; 2000-602362/57.
XX DR N-PSDB; AAC75154.
XX PS
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease.
XX PS
XX PS Claim 11; Page 1203-1206; 5507pp; English.
XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX CC antiproliferative; antiparkinsonian; neurotrophic; immunostimulant;
XX CC anticarcinogenic; antiarthritic; immunosuppressive; osteopathic;
XX CC cardiolipid; thrombolytic; coagulant; vasotrophic; antidiabetic; hypotensive;
XX CC dermatological; immunosuppressive; antineoplastic; antibacterial;
XX CC antiviral; antifungal; antineuritic; antithyroid; and antianemic. The
XX CC sequences can be used for determining the presence of or predisposition
XX CC to, or preventing or treating pathological conditions associated with an
XX CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX CC proteins in gene therapy vectors. The proteins and nucleic acids may be
XX CC used to treat cancers, proliferative disorders, neurodegenerative
XX CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX CC storage, systemic lupus erythematosus, severe combined immunodeficiency
XX CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
XX CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX SQ Sequence 1532 AA;

Query Match 1.5%; Score 8; DB 3; Length 1532;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 513 STRASGQT 520
DB 783 STRASGQT 790

RESULT 48
AAB82820
ID AAB82820 standard; peptide; 7 AA.
XX AC AAB82820;
XX DT 12-NOV-2001 (first entry)
XX DE Human low density lipoprotein binding protein 3 (LBP-3) peptide.
XX KW Low density lipoprotein binding protein 3; LBP-3; LDL; human;
XX KW atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.
XX OS Homo sapiens.
XX FN WO200164874-A2.
XX PD 07-SEP-2001.
XX PF 28-FEB-2001; 2001WO-US006356.
XX PR 02-MAR-2000; 2000US-00517849.
XX PR 14-JUL-2000; 2000US-00616289.
XX PA (BOST-) BOSTON HEART FOUND INC.
XX PI Lees AM, Lees RS, Law SW, Arjona AA;

Query Match 1.3%; Score 7; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 RDVSEEL 59
DB 1 RDVSEEL 7

RESULT 49
AAB50054
ID AAB50054 standard; peptide; 15 AA.
XX AC AAB50054;
XX DT 26-JUL-2002 (first entry)
XX DE Proline-enriched gamma-carboxyl glutamate-protein 194.05 N-terminal.
XX KW Proline-enriched gamma-carboxyl glutamate-protein 194.05; cancer;
XX KW inflammation.
XX OS Unidentified.
XX PN CNI331106-A.
XX PD 16-JAN-2002.
XX PF 30-JUN-2000; 2000CN-00116909.
XX PR 30-JUN-2000; 2000CN-00116909.
XX PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX PI Mao Y, Xie Y;
XX PD WPI; 2002-316393/36.
XX PT Proline-enriched gamma-carboxyl glutamate-protein 194.05 polypeptide and
XX PT its encoding polynucleotide, for treating e.g. cancer and inflammation.
XX PS Example 5; Page 20 (Disclosure); 37pp; Chinese.
XX CC This invention describes a novel proline-enriched gamma-carboxyl
XX CC glutamate-protein 194.05 polypeptide and its encoding nucleic acid. The
XX CC polynucleotide, polypeptide and its antagonist are useful for treating

CC e.g. cancer and inflammation. This sequence represents the proline-
 CC enriched gamma-carboxyl glutamate protein 194.05 N-terminal fragment
 CC described in the method of the invention
 XX
 XX Sequence 15 AA;
 SQ
 Query Match 1.3%; Score 7; DB 5; Length 15;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 GQPEAGP 11
 DB 6 GQPEAGP 12
 RESULT 50
 ADD24103
 ID ADD24103 standard; peptide; 15 AA.
 XX
 AC ADD24103;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Breast cancer membrane protein (BCMP) peptide.
 XX
 KW breast cancer; screening; diagnosis; breast cancer therapy;
 KW breast cancer membrane protein; BCMP; cytostatic; vaccine; human.
 XX
 OS Homo sapiens.
 XX
 PN WO2003087831-A2.
 XX
 PD 23-OCT-2003.
 XX
 PF 10-APR-2003; 2003WO-GB001559.
 XX
 PR 11-APR-2002; 2002GB-00008331.
 XX
 PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
 XX
 PI Hudson LJ, Stamps AC, Terrett JA;
 XX
 DR WPI; 2003-845381/78.
 XX
 PT Screening, diagnosing and/or treating breast cancer by detecting a change
 PT in expression or activity of a breast cancer membrane protein (BCMP)
 PT polypeptide or encoding nucleic acid molecule.
 XX
 PS Example; Page 75; B1pp; English.
 XX
 CC The present invention describes a method of screening for and/or
 CC diagnosing breast cancer in a subject, and/or monitoring the
 CC effectiveness of breast cancer therapy. The method comprises detecting
 CC and/or quantifying in a biological sample obtained from the subject a
 CC breast cancer membrane protein (BCMP) polypeptide and a nucleic acid
 CC molecule. Also described: (1) an antibody, its functionally-active
 CC fragment, derivative or analogue, that specifically binds to one or more
 CC of the BCMP polypeptide; (2) a diagnostic kit comprising a capture
 CC reagent specific for an BCMP polypeptide, reagents and instructions for
 CC use; (3) a method for screening for anti-breast cancer agents that
 CC interact with the BCMP polypeptide, comprising contacting the polypeptide
 CC with a candidate agent, and determining whether or not the candidate
 CC agent interacts with the polypeptide; (4) a method for screening for anti-
 CC breast cancer agents that modulate the expression or activity of an BCMP
 CC polypeptide or the nucleic acid molecule cited above, comprising
 CC comparing the expression or activity of the polypeptide or nucleic acid
 CC molecule, in the presence and absence of a candidate agent or in the
 CC presence of a control agent, and determining whether the candidate agent
 CC causes the expression or activity of the polypeptide or nucleic acid
 CC molecule to change; and (5) an agent identified by the method of (3) or
 CC (4), which interacts with the polypeptide or causes the expression or
 CC activity of the polypeptide, or the expression of the nucleic acid
 CC molecule to change. BCMPs have cytostatic activities, and can be used in

CC vaccines. The BCMP polypeptide, nucleic acid molecule, antibody, agent or
 CC their derivatives, are useful in the manufacture of a medicament for the
 CC treatment of breast cancer, where the composition is a vaccine. The
 CC present sequence represents a BCMP peptide which is used in the
 CC exemplification of the present invention. N.B. The present sequence is
 CC designated as SEQ ID NO:1054 on page 75, but does not correspond with the
 CC previously given SEQ ID NO:1054 in the specification.
 XX
 SQ Sequence 15 AA;
 Query Match 1.3%; Score 7; DB 7; Length 15;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 473 SAGGQGS 479
 DB 3 SAGGQGS 9
 RESULT 51
 ABP83398
 ID ABP83398 standard; peptide; 18 AA.
 XX
 AC ABP83398;
 XX
 DT 04-MAR-2003 (first entry)
 XX
 DE G protein-coupled receptor (GPCR) antigenic peptide SEQ ID NO:2071.
 XX
 KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
 KW G protein-coupled receptor modulator; antibody; immune-related disease;
 KW growth-related disease; cell regeneration-related disease; AIDS; cancer;
 KW immunological-related cell proliferative disease; autoimmune disease;
 KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
 KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
 KW graft-versus host disease; Parkinson's disease; multiple sclerosis; pain;
 KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
 KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
 KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
 KW ulcer.
 XX
 OS Homo sapiens.
 XX
 PN WO200261087-A2.
 XX
 PD 08-AUG-2002.
 XX
 PF 19-DEC-2001; 2001WO-US050107.
 XX
 PR 19-DEC-2000; 2000US-0257144P.
 XX
 PA (LIFE-) LIFESPAN BIOSCIENCES INC.
 XX
 PI Burner GC, Roush CL, Brown JP;
 XX
 DR WPI; 2003-046718/04.
 XX
 XX New isolated antigenic peptides e.g., for G protein-coupled receptors
 PT (GPCR), useful for diagnosing and designing drugs for treating conditions
 PT in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or
 PT autoimmune diseases.
 XX
 PS Claim 1; Fig 2; 523pp; English.
 XX
 CC The present invention describes antigenic peptides (I) comprising: (a)
 CC any one of 1601 sequences (see ABP82019 to ABP82619) of 12-24 amino
 CC acids. Also described: (1) an assay for the detection of a particular G
 CC protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
 CC and (2) an isolated antibody having high specificity and high affinity or
 CC avidity for a particular GPCR. (I) can be used as GPCR modulators and in
 CC gene therapy. The antigenic peptides for GPCRs are useful in detecting an
 CC antibody against a particular GPCR, and in the production of specific
 CC antibodies. The peptides and antibodies are also useful for detecting the

presence or absence of corresponding GPCRs. The antigenic peptides for GPCRs and antibodies are useful for diagnosing and designing drugs for treating immune-related diseases, growth-related diseases, cell regeneration-related disease, immunological-related cell proliferative diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease, atherosclerosis, bacterial, fungal, protozoan or viral infections, osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic acute inflammation, allergies, Crohn's disease, diabetes, graft versus host disease, Parkinson's disease, multiple sclerosis, pain, psoriasis, anxiety, depression, schizophrenia, dementia, mental retardation, memory loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension, hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or any other disorder in which GPCRs are involved. The antibodies may be used in immunoassays and immunodiagnosis. A2Z42523 to A2Z42869 encode GPCR proteins given in ABP91675 to ABP92018, which are used in the exemplification of the present invention.

Sequence 18 AA;

Query Match	1.3%;	Score 7;	DB 6;	Length 18;
Best Local Similarity	100.0%;	Pred. No. 75;		
Matches 7;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	57	EELSRQL	63
Dd	6	EELSRQL	12

RESULT 52
AAW03444
ID AAW03444 standard; peptide; 25 AA.

AAW03444;

DT 14-FEB-1997 (first entry)

XX HDV antigenic peptide, delta25-49.

XX Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
KW coiled-coil dimerisation domain; epitope; multimer.

XX Synthetic.

XX
PN
W09620953-A2.XX
11-1111-1996

XX
DE
22-DEC-1995.
95W0-115016854

XX
30-DEC-1994. 0941S-00366479

XX
DA
(TIC -) INTY NORPU CAROLINA

XX
XX
XX

XX

XX
PT
Synthetic peptide with Delta hepatitis virus antigenic activity - useful

XXXX

xx The sequences given in AA003442-47 represent peptides derived from the
CC hepatitis delta virus (HDV) antigen (HDAg) sequence. These sequences
CC represent synthetic peptides with hepatitis delta virus (HDV) antigenic
CC activity. They were used to raise murine monoclonal antibodies which may
CC be used to immunise animals against HDV, and to detect anti-HDV
CC antibodies. This sequence corresponds to segment B of HDAg
CC

Sequence 25 AA:

Query Match 1.3%; Score 7; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 99;

	Matches	7; Conservative	Mismatches	0; Indels	0; Gaps
QY	404	KKIKKLE 410			
Db	15	KKIKKLE 21			

RESULT 53
AA64949
ID AA64949 standard; protein; 27 AA.

AA
AC AAY64949;

XX	
DT	01-FEB-2000 (first entry)

Human 5' EST related polypeptide SEQ ID NO:1110.

Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
 gene therapy; chromosome mapping; upstream regulatory sequence; forensic;
 location; development; protein synthesis; stability; regulation;
 identification.

XX OS Homo sapiens.

XX PN W09953051-A2

XX
21-0000-1999

XX DE 09-APP-1000. 00W0-7B000712

XX
09-APR-1999. 08115-00057719

PR 28-APR-1998; 98US-00069047.
vv

PA (GEST) GENSET .

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-038446/03.



PT diagnostic, forensic, gene therapy, and chromosome mapping procedures.

PS Claim 3; Page 689; 837pp; English.

AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST) sequences, corresponding to human secreted proteins. AAY64651 to AAY65438 represent the EST-related proteins corresponding to AAZ242965 to AAZ43052. The 5' ESTs can be used for producing secreted human gene products. They can be used to identify and isolate 5' untranslated regions (UTRs) and upstream regulatory regions which control the location, development stage, rate, and quantity of protein synthesis, as well as stability of mRNA. The ESTs are also useful as probes for chromosome mapping, and to obtain full length cDNA clones. The ESTs can also be used in forensic procedures to identify individuals, or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal gene expression. The products may also be used in gene therapy protocols. The nucleic acids encoding signal peptides can be used for directing extracellular secretion of a polypeptide or the insertion of a polypeptide into a membrane, or importing a polypeptide into a cell. The proteins encoded by the EST sequences may be useful in treating a variety of human conditions. Secreted proteins have therapeutic value, and the identification of new secreted proteins is valuable. AAZ42449 to AAZ42264 and AAY64644 to AAY64650 represent sequences used in the exemplification of the present invention.

XX
SO
Sequence 27 AA;

```

Query Match      1.3%; Score 7; DB 3; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

047 493 GACAPSS 499

DB |||||
4 GAQAPSS 10

RESULT 54
AAW03445
ID AAW03445 standard; peptide; 32 AA.
XX AC AAW03445;
XX DT 14-FEB-1997 (first entry)
XX HDV antigenic peptide, delta18-49.
XX Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX coiled-coil dimerisation domain; epitope; multimer.
XX Synthetic.
XX Key Location/Qualifiers
FT Peptide 8..32
FT /label= Segment B
FT /note= "Corresponds to HDAG residues 25-49"
XX W09620953-A2.
XX 11-JUL-1996.
XX 22-DEC-1995; 95WO-US016854.
XX 30-DEC-1994; 94US-00366479.
XX (TYNC-) UNIV NORTH CAROLINA.
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX WPI; 1996-333940/33.
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX to produce and detect viral antibodies in a mammal and for immunisation.
XX Example 10; Page 29; 50pp; English.
XX The sequences given in AAW03442-47 represent peptides derived from the
XX hepatitis delta virus (HDV) antigen (HDAG) sequence. These sequences
XX represent synthetic peptides with hepatitis delta virus (HDV) antigenic
XX activity. They were used to raise murine monoclonal antibodies which may
XX be used to immunise animals against HDV, and to detect anti-HDV
XX antibodies. This sequence corresponds to most of segment A and all of
XX segment B of HDAG
XX Sequence 32 AA;
Query Match 1.3%; Score 7; DB 2; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KXIKKLE 410
DB 22 KXIKKLE 28

RESULT 55
ABP97458
ID ABP97458 standard; peptide; 34 AA.
XX AC ABP97458;
XX DT 04-AUG-2003 (first entry)
XX HIV N36-binding peptide SC34-b.
XX HIV; human immunodeficiency virus; N36-binding peptide; drug composite;

KW reverse transcriptase inhibitor; HIV protease inhibitor;
KW chemokine receptor antagonist; drug targeting; AIDS;
KW acquired immunodeficiency syndrome;
KW highly active anti-retroviral therapy; HAAR; anti-HIV.
XX Synthetic.
XX Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-terminal acetyl"
FT Modified-site 34
FT /note= "C-terminal amide"
XX W02003029284-A1.
XX 10-APR-2003.
XX 27-SEP-2002; 2002WO-JP010119.
XX 27-SEP-2001; 2001JP-00297963.
XX (FUJI) FUJII N.
XX Fujii N, Otaka A, Matsuoka M;
XX WPI; 2003-333562/31.
XX Anti-HIV agents containing polypeptide with high affinity toward N36 of
XX human immunodeficiency virus, or/and its composites, in drug compositions
XX for preventing or treating AIDS or onset of AIDS.
XX Example 1 #2; Page 11; 44pp; Japanese.
XX The invention relates to peptides with the ability to bind to N36 of
XX human immunodeficiency virus (HIV). The peptides have high affinity for
XX HIV N36, and have potent anti-HIV activity and high water solubility. The
XX peptides may optionally be linked to a reverse transcriptase inhibitor,
XX HIV protease inhibitor or chemokine receptor antagonist to form a
XX composite, and may also be used to target drugs to HIV-infected cells.
XX Peptides and/or composites of the invention may be used to prevent or
XX treat the onset of AIDS in HIV positive individuals or AIDS patients,
XX either alone or in combination with AIDS highly active anti-retroviral
XX therapies (HAAR). Sequences ABP97457-ABP97461 represent N36-binding
XX peptides of the invention
XX Sequence 34 AA;
Query Match 1.3%; Score 7; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 403 TKKIKKL 409
DB 12 TKKIKKL 18

RESULT 56
ABP97457
ID ABP97457 standard; peptide; 34 AA.
XX AC ABP97457;
XX DT 04-AUG-2003 (first entry)
XX HIV N36-binding peptide SC34-a.
XX HIV; human immunodeficiency virus; N36-binding peptide; drug composite;
KW reverse transcriptase inhibitor; HIV protease inhibitor;
KW chemokine receptor antagonist; drug targeting; AIDS;
KW acquired immunodeficiency syndrome;
KW highly active anti-retroviral therapy; HAAR; anti-HIV.
XX Synthetic.

```

XX Key Location/Qualifiers
FT Modified-site 1 /note= "The N-terminus is either acetylated, or is linked
FT via a CO-(CH2)3-CO-NH-(CH2)5-CO- linker to AZT"
FT Modified-site 2
FT Modified-site 34 /label= Nle
FT Modified-site 34 /note= "C-terminal amide"
XX
XX WO2003029284-A1.
XX
XX 10-APR-2003.
XX
XX 27-SEP-2002; 2002WO-JP010119.
XX
XX 27-SEP-2001; 2001JP-00297963.
XX
XX (FUJII) FUJII N.
XX
XX Fujii N, Otake A, Matsuo M;
XX
XX WPI; 2003-333562/31.
XX
XX Anti-HIV agents containing polypeptide with high affinity toward N36 of
XX human immunodeficiency virus, or/and its composites, in drug compositions
XX for preventing or treating AIDS or onset of AIDS.
XX
XX Example 1 #1; Page 11; 44pp; Japanese.
XX
XX The invention relates to peptides with the ability to bind to N36 of
XX human immunodeficiency virus (HIV). The peptides have high affinity for
XX HIV N36, and have potent anti-HIV activity and high water solubility. The
XX peptides may optionally be linked to a reverse transcriptase inhibitor,
XX HIV protease inhibitor or chemokine receptor antagonist to form a
XX composite, and may also be used to target drugs to HIV-infected cells.
XX Peptides and/or composites of the invention may be used to prevent or
XX treat the onset of AIDS in HIV positive individuals or AIDS patients,
XX either alone or in combination with AIDS highly active anti-retroviral
XX therapies (HAAR). Sequences ABP97457-ABP97461 represent N36-binding
XX peptides of the invention
XX
XX Sequence 34 AA;
XX
XX Query Match 1.3%; Score 7; DB 6; Length 34;
XX Best Local Similarity 100.0%; Pred. No. 1.3e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 403 TKIKKL 409
XX |||||
XX 12 TKIKKL 18
XX
XX
XX RESULT 57
XX AAW03447
XX ID AAW03447 standard; peptide; 35 AA.
XX
XX AC AAW03447;
XX
XX 14-FEB-1997 (first entry)
XX
XX HDV antigenic peptide, delta(C)28-60(Y).
XX
XX Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX coiled-coil dimerisation domain; epitope; multimer.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Peptide 1.23
XX /label= Segment B
XX /note= "Corresponds to HDag residues Cys-28-49"
XX
XX Peptide 24.34

```

```

FT /label= Segment_C
FT /note= "Corresponds to HDag residues 50-60"
XX
XX WO9620953-A2.
XX
XX 11-JUL-1996.
XX
XX 22-DEC-1995; 95WO-US016854.
XX
XX 30-DEC-1994; 94US-00366479.
XX
XX (UYNC-) UNIV NORTH CAROLINA.
XX
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX WPI; 1996-333940/33.
XX
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX to produce and detect viral antibodies in a mammal and for immunisation.
XX
XX Example 10; Page 29; 50pp; English.
XX
XX The sequences given in AAW03442-47 represent peptides derived from the
XX hepatitis delta virus (HDV) antigen (HDag) sequence. These sequences
XX represent synthetic peptides with hepatitis delta virus (HDV) antigenic
XX activity. They were used to raise murine monoclonal antibodies which may
XX be used to immunise animals against HDV, and to detect anti-HDV
XX antibodies. This sequence corresponds to most of segment B and all of
XX segment C of HDag
XX
XX Sequence 35 AA;
XX
XX Query Match 1.3%; Score 7; DB 2; Length 35;
XX Best Local Similarity 100.0%; Pred. No. 1.3e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 404 KXIKLE 410
XX |||||
XX 13 KXIKLE 19
XX
XX
XX RESULT 58
XX AAW03446
XX ID AAW03446 standard; peptide; 35 AA.
XX
XX AC AAW03446;
XX
XX 14-FEB-1997 (first entry)
XX
XX HDV antigenic peptide, delta15-49.
XX
XX Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX coiled-coil dimerisation domain; epitope; multimer.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Peptide 11.35
XX /label= Segment_B
XX /note= "Corresponds to HDag residues 25-49"
XX
XX WO9620953-A2.
XX
XX 11-JUL-1996.
XX
XX 22-DEC-1995; 95WO-US016854.
XX
XX 30-DEC-1994; 94US-00366479.
XX
XX (UYNC-) UNIV NORTH CAROLINA.
XX
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX
XX

```

DR WPI; 1996-333940/33.
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
PT to produce and detect viral antibodies in a mammal and for immunisation.
XX
PS Example 10; Page 29; 50pp; English.
XX
CC The sequences given in AA03442-47 represent peptides derived from the
CC hepatitis delta virus (HDV) antigen (HDAG) sequence. These sequences
CC represent synthetic peptides with hepatitis delta virus (HDV) antigenic
CC activity. They were used to raise murine monoclonal antibodies which may
CC be used to immunise animals against HDV, and to detect anti-HDV
CC antibodies. This sequence corresponds to most of segment A and all of
CC segment B of HDAG
XX
SQ Sequence 35 AA;
Query Match 1.3%; Score 7; DB 2; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KRIKLE 410
DB 25 KRIKLE 31
RESULT 59
ABG23756
ID ABG23756 standard; protein; 36 AA.
XX
AC ABG23756;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #23747.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS87943.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
PS Claim 20; SEQ ID NO 54115; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (I) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC

CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG0377 represent novel human diagnostic
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 36 AA;
Query Match 1.3%; Score 7; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 477 QGSLTDS 483
DB 19 QGSLTDS 25
RESULT 60
ABG27966
ID ABG27966 standard; protein; 36 AA.
XX
AC ABG27966;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27957.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS92153.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
PS Claim 20; SEQ ID NO 58325; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (I) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC


```

ABG27959
ID  ABG27959 standard; protein; 37 AA.
XX
AC  ABG27959;
XX
DT  18-FEB-2002 (first entry)
XX
DE  Novel human diagnostic protein #27950.
XX
XX  Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX  food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS  Homo sapiens.
XX
XX  WO200175067-A2.
XX
XX  11-OCT-2001.
XX
XX  30-MAR-2001; 2001WO-US008631.
XX
XX  31-MAR-2000; 2000US-00540217.
XX
XX  23-AUG-2000; 2000US-00649167.
XX
XX  (HYSE-) HYSEQ INC.
XX
XX  Drmanac RT, Liu C, Tang YT;
XX
XX  WPI: 2001-639362/73.
XX
XX  N-PSDB; AAS92146.
XX
XX  New isolated polynucleotide and encoded polypeptides, useful in
XX  diagnostics, forensics, gene mapping, identification of mutations
XX  responsible for genetic disorders or other traits and to assess
XX  biodiversity.
XX
XX  Claim 20; SEQ ID NO 58318; 103pp; English.
XX
XX  The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX  sequences. (I) is useful as hybridisation probes, polymerase chain
XX  reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX  and in recombinant production of (II). The polynucleotides are also used
XX  in diagnostics as expressed sequence tags for identifying expressed
XX  genes. (I) is useful in gene therapy techniques to restore normal
XX  activity of (II) or to treat disease states involving (II). (II) is
XX  useful for generating antibodies against it, detecting or quantitating a
XX  polypeptide in tissue, as molecular weight markers and as a food
XX  supplement. (II) and its binding partners are useful in medical imaging
XX  of sites expressing (II). (I) and (II) are useful for treating disorders
XX  involving aberrant protein expression or biological activity. The
XX  polypeptide and polynucleotide sequences have applications in
XX  diagnostics, forensics, gene mapping, identification of mutations
XX  responsible for genetic disorders or other traits to assess biodiversity
XX  and to produce other types of data and products dependent on DNA and
XX  amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
XX  amino acid sequences of the invention. Note: The sequence data for this
XX  patent did not appear in the printed specification, but was obtained in
XX  electronic format directly from WIPO at
XX  ftp.wipo.int/pub/published_pcr_sequences
XX
XX  Sequence 37 AA;
XX
XX  Query Match 1.3%; Score 7; DB 4; Length 37;
XX  Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  477 QGSLTDS 483
DB  20 QGSLTDS 26
|||||

RESULT 64
AAW03442
ID  AAW03442 standard; peptide; 38 AA.
XX
AC  AAW03442;
XX
DT  14-FEB-1997 (first entry)
XX
DE  HDV antigenic peptide, delta12-49.
XX
XX  Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX  coiled-coil dimerisation domain; epitope; multimer.
XX
OS  Synthetic.
XX
XX  WO9620953-A2.
XX
XX  11-JUL-1996.
XX
XX  22-DEC-1995; 95WO-US016854.
XX
XX  30-DEC-1994; 94US-00366479.
XX
XX  (UTNC-) UNIV NORTH CAROLINA.
XX
XX  Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX
XX  WPI: 1996-333940/33.
XX
XX  Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX  to produce and detect viral antibodies in a mammal and for immunisation.
XX
XX  Example 10; Page 29; 50pp; English.
XX
XX  The sequences given in AAW03442-47 represent peptides derived from the
XX  hepatitis delta virus (HDV) antigen (HDAG) sequence. These sequences
XX  represent synthetic peptides with hepatitis delta virus (HDV) antigenic
XX  activity. They were used to raise murine monoclonal antibodies which may
XX  be used to immunise animals against HDV, and to detect anti-HDV
XX  antibodies
XX
XX  Sequence 38 AA;
XX
XX  Query Match 1.3%; Score 7; DB 2; Length 38;
XX  Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  404 KKIKKLE 410
DB  28 KKIKKLE 34
|||||

RESULT 65
AAW16323
ID  AAW16323 standard; protein; 38 AA.
XX
XX  AAW16323;
XX
XX  12-OCT-2001 (first entry)
XX
XX  Peptide #2757 encoded by probe for measuring cervical gene expression.
XX
XX  Probe; human; microarray; gene expression; cervical epithelial cell;
XX  cervical cancer.
XX
XX  Homo sapiens.
XX
XX  WO200157278-A2.
XX
XX  09-AUG-2001.
XX
XX  30-JAN-2001; 2001WO-US000670.
XX
XX  04-FEB-2000; 2000US-0180312P.
XX
XX  26-MAY-2000; 2000US-0207456P.
XX
XX  30-JUN-2000; 2000US-00608408.

```

```

PT 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
PS 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX Claim 27; SEQ ID NO 21149; 487pp; English.
XX The present invention relates to human single exon nucleic acid probes
CC (SENP; see AA110069-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 38 AA;
XX
XX Query Match 1.3%; Score 7; DB 4; Length 38;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 281 KLRQENM 287
XX |
XX 23 KLRQENM 29
XX |
XX
XX RESULT 66
XX ABB35315
XX ID ABB35315 standard; peptide; 38 AA.
XX AC ABB35315;
XX
XX DT 04-FEB-2002 (first entry)
XX
XX DE Peptide #2821 encoded by human foetal liver single exon probe.
XX
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
XX OS Homo sapiens.
XX
XX PN WO200157277-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000669.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT

```

```

PT gene expression in human fetal liver.
XX Claim 27; SEQ ID NO 27950; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 38 AA;
XX
XX Query Match 1.3%; Score 7; DB 4; Length 38;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 281 KLRQENM 287
XX |
XX 23 KLRQENM 29
XX |
XX
XX RESULT 67
XX AAM28818
XX ID AAM28818 standard; protein; 38 AA.
XX AC AAM28818;
XX
XX DT 17-OCT-2001 (first entry)
XX
XX DE Peptide #2855 encoded by probe for measuring placental gene expression.
XX
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO200157272-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000663.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488997/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human placenta.
XX Claim 27; SEQ ID NO 23087; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP;
CC see AA131315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders
XX Sequence 38 AA;
XX

```

Query Match 1.3%; Score 7; DB 4; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 KLRQENM 287
|||
Db 23 KLRQENM 29

RESULT 68
ABB30145
ID ABB30145 standard; peptide; 38 AA.

XX AC ABB30145;
XX 01-FEB-2002 (first entry)
XX DE Peptide #2796 encoded by breast cell single exon nucleic acid probe.
XX Human; microarray; single exon probe; gene expression; breast; disease;
XX cancer.
XX Homo sapiens.
XX PN WO200157271-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000662.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX DR

XX PT New spatially-addressable set of single exon nucleic acid probes, useful
XX for measuring gene expression in sample derived from human breast,
XX comprises number of single exon nucleic acid probes.
XX PS Claim 27; SEQ ID NO 13113; 327pp + Sequence Listing; English.
XX CC The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting the
XX probes with a collection of detectably labelled nucleic acids derived
XX from mRNA of human breast, and then measuring the label bound to each
XX probe of the microarray. The probes are useful for verifying the
XX expression of regions of genomic DNA predicted to encode proteins. They
XX are useful for gene discovery, and for determining predisposition and/or
XX assessing the toxicity of chemical agents on cells. The microarray of
XX this invention presents a far greater diversity of probes for measuring
XX gene expression, with far less bias than expressed sequence tag
XX microarrays. The method is suitable for rapid production of functional
XX information from genomic sequence. The present sequence is a peptide
XX encoded by a single exon nucleic acid probe of the invention. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 38 AA;

Query Match 1.3%; Score 7; DB 4; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 KLRQENM 287
|||
Db 23 KLRQENM 29

RESULT 69
ABB20757
ID ABB20757 standard; protein; 38 AA.

XX AC ABB20757;
XX 23-JAN-2002 (first entry)
XX DE Protein #2756 encoded by probe for measuring heart cell gene expression.
XX Human; gene expression; heart; microarray; vascular system;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease.
XX OS Homo sapiens.
XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000666.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX DR

XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
XX PS Claim 15; SEQ ID NO 22527; 530pp; English.
XX CC The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting,
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease.
XX hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 38 AA;

Query Match 1.3%; Score 7; DB 4; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 KLRQENM 287
|||
Db 23 KLRQENM 29

RESULT 70

XX PD 09-AUG-2001.
 XX XX 30-JAN-2001; 2001WO-US000667.
 XX PF 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX PD WO200157276-A2.
 XX PF 09-AUG-2001.
 XX PR 30-JAN-2001; 2001WO-US000668.
 XX PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX PD WO200157276-A2.
 XX PF 09-AUG-2001.
 XX PR 30-JAN-2001; 2001WO-US000667.
 XX PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX PD WO200157276-A2.
 XX PF 09-AUG-2001.
 XX PR 30-JAN-2001; 2001WO-US000664.
 XX PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.

AAM68519
 ID AAM68519 standard; protein; 38 AA.
 AC AAM68519;
 XX 06-NOV-2001 (first entry)
 DT Human bone marrow expressed probe encoded protein SEQ ID NO: 28825.
 DE Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma.
 XX OS Homo sapiens.
 XX PD WO200157276-A2.
 XX PF 09-AUG-2001.
 XX PR 30-JAN-2001; 2001WO-US000668.
 XX PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX PD WO200157276-A2.
 XX PF 09-AUG-2001.
 XX PR 30-JAN-2001; 2001WO-US000668.
 XX PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX PD WO200157276-A2.
 XX PF 09-AUG-2001.
 XX PR 30-JAN-2001; 2001WO-US000667.
 XX PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.

AAM56146
 ID AAM56146 standard; protein; 38 AA.
 AC AAM56146;
 XX 05-NOV-2001 (first entry)
 DT Human brain expressed single exon probe encoded protein SEQ ID NO: 28251.
 DE Human; brain expressed exon; gene expression analysis; probe; microarray;
 KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
 XX OS Homo sapiens.
 XX PD WO200157275-A2.
 XX PF 05-NOV-2001 (first entry)
 DT Human brain expressed single exon probe encoded protein SEQ ID NO: 28251.
 DE Human; brain expressed exon; gene expression analysis; probe; microarray;
 KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
 XX OS Homo sapiens.
 XX PD WO200157275-A2.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48898/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX Claim 27; SEQ ID NO 28836; 658bp; English.
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridizes at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.
XX (I) may be used for predicting, measuring and displaying gene expression
XX in samples derived from human adult liver. The genes identified may be
XX involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX associated with coronary heart disease. ABG47348-ABG59930 represent human
XX liver single exon encoded peptides of the invention. Note: The sequence
XX information for this patent does not appear in the printed specification
XX but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 38 AA;
XX
XX Query Match 1.3%; Score 7; DB 4; Length 38;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 281 KLRQENM 287
XX DB 23 KLRQENM 29
XX
XX RESULT 73
XX AA04061
XX ID AA04061 standard; protein; 38 AA.
XX AC AA04061;
XX DT 09-OCT-2001 (first entry)
XX DE Peptide #2743 encoded by probe for measuring breast gene expression.
XX KW Probe; human; breast disease; breast cancer; development disorder;
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN WO200157270-A2.
XX PD 09-AUG-2001.
XX PF 29-JAN-2001; 2001WO-US000661.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression in
XX a human breast.

XX PS Claim 27; SEQ ID NO 12801; 322pp; English.
XX The present invention relates to novel single exon nucleic acid probes
XX (see AA00010-AA110067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridizes at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for
XX predicting, diagnosing, grading, staging, monitoring and prognosing
XX diseases of the human breast, particularly those diseases with polygenic
XX aetiology. The diseases include: breast cancer, disorders of development,
XX inflammatory diseases of the breast, fibrocystic changes, proliferative
XX breast disease and non-carcinoma tumours. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 38 AA;
XX
XX Query Match 1.3%; Score 7; DB 4; Length 38;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 281 KLRQENM 287
XX DB 23 KLRQENM 29
XX
XX RESULT 74
XX ABG38100
XX ID ABG38100 standard; peptide; 38 AA.
XX AC ABG38100;
XX DT 19-AUG-2002 (first entry)
XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 27765.
XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX KW chronic obstructive pulmonary disease; interstitial lung disease;
XX KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX KW primary ciliary dyskinesia; pulmonary hypertension;
XX KW hyaline membrane disease.
XX OS Homo sapiens.
XX PN WO200186003-A2.
XX PD 15-NOV-2001.
XX PF 30-JAN-2001; 2001WO-US000665.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples.

PS Claim 27; SEQ ID NO 27765; 634pp; English.

XX The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived

CC from human lung comprising single exon nucleic acid probes having one of

CC 12614 nucleic acid sequences mentioned in the specification, or their

CC complements or the 12387 open reading frames derived from the 12614

CC probes. Also included are a microarray comprising the novel set of probes

CC; the novel set of probes which hybridise at high stringency to a nucleic

CC acid expressed in the human lung; measuring gene expression in a sample

CC derived from human lung, comprising (a) contacting the array with a

CC collection of detectably labeled nucleic acids derived from human lung

CC mRNA; and (b) measuring the label detectably bound to each probe of the

CC array; identifying exons in a eukaryotic genome, comprising (a)

CC algorithmically predicting at least one exon from genomic sequences of

CC the eukaryote; and (b) detecting specific hybridisation of detectably

CC labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,

CC having a fragment identical to the predicted exon, the probe is included

CC in the above mentioned microarray; assigning exons to a single gene,

CC comprising (a) identifying exons from genomic sequence by the method

CC above and (b) measuring the expression of each of the exons in several

CC tissues and/or cell types using hybridisation to a single exon

CC microarrays having a probe with the exon, where a common pattern of

CC expression of the exons in the tissues and/or cell types indicates that

CC the exons should be assigned to a single gene; a peptide comprising one

CC of 12011 sequences, mentioned in the specification, or encoded by the

CC probes/open reading frames (ORF). The probes are used for gene expression

CC analysis, and for identifying exons in a gene, particularly using human

CC lung derived mRNA and for the study of lung diseases such as asthma, lung

CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung

CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,

CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-

CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary

CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,

CC Karagenar syndrome, fibrocystic pulmonary dysplasia, primary ciliary

CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The

CC present sequence is a peptide/protein encoded by a single exon probe of

CC the invention. Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic format

CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 33 AA;

Query Match 1.3%; Score 7; DB 5; Length 38;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 KLRQNM 287

DB 23 KLRQNM 29

|||||

RESULT 75

ABG28175

ID ABG28175 standard; protein; 42 AA.

XX

AC ABG28175;

XX

DF 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #28166.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

XX

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

DR N-PSDB; AAS92362.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

XX Claim 20; SEQ ID NO 58534; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is

CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic

CC amino acid sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 42 AA;

Query Match 1.3%; Score 7; DB 4; Length 42;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 QGSLTDS 483

DB 25 QGSLTDS 31

|||||

RESULT 76

ABG23760

ID ABG23760 standard; protein; 43 AA.

XX

AC ABG23760;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #23751.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR N-PSDB; AAS87947.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 54119; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 43 AA;
SQ

Query Match 1.3%; Score 7; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 QGSLTDS 483
DB 26 QGSLTDS 32
|||||

RESULT 77
ABG23761
ID ABG23761 standard; protein; 43 AA.
XX
XX AC ABG23761;
XX
XX DT 18-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #23752.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensics;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US008631.
XX
XX PR 31-MAR-2000; 2000US-00540217.
XX
XX PR 23-AUG-2000; 2000US-00649167.
XX
XX PA (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS87948.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 54120; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 43 AA;
SQ

Query Match 1.3%; Score 7; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 QGSLTDS 483
DB 26 QGSLTDS 32
|||||

RESULT 78
AA95964
ID AA95964 standard; protein; 44 AA.
XX
XX AC AA95964;
XX
XX DT 20-NOV-2000 (first entry)
XX
XX DE Construct used in post-translational modification assay.
XX
XX KW Coiled-coil; post-translational modification; protease; assay; assay;
XX signal modulation.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 1 /note= "amino acid used for measuring reporter output"
XX FT Peptide 2..15 /label= Leader_Peptide
XX FT Misc-difference 16..19 /note= "GPI attachment site"
XX FT Peptide 20..44 /note= "recognition sequence; residues 20-29 are
XX hydrophilic and residues 30-44 are hydrophobic"
XX
XX PN WO200050635-A1.

XX 31-AUG-2000.
 PD 18-FEB-2000; 2000WO-GB000601.
 XX 25-FEB-1999; 99GB-00004401.
 XX (FLUO-) FLUORESCENCE LTD.
 XX Coyle J, Lightowler J;
 XX WPI; 2000-565459/52.
 DR Labeled polypeptide binding partner compositions useful for monitoring
 XX protease activity by detecting signal modulation.
 XX Example 6; Page 63; 90pp; English.
 PS The present sequence is that of a polypeptide that adopts a
 CC heteromultimeric coiled-coil conformation. It is used in an example of
 CC the invention to measure post-translational modification events which
 CC have proteolysis as an integral step. A polypeptide useful in the present
 CC invention is capable of multimerising with similar or different
 CC polypeptides. The multimer comprises a first polypeptide associated with
 CC a label and a second polypeptide, where (a) at least 1 of the
 CC polypeptides is susceptible to protease digestion, (b) association of the
 CC polypeptides to form a multimer is detectable via a signal emitted by the
 CC signal, and (c) digestion of at least 1 polypeptides results in
 CC dissociation of the multimer and modulation of the signal emitted by the
 CC label
 XX Sequence 44 AA;
 SQ
 Query Match 1.3%; Score 7; DB 3; Length 44;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 406 IKLKEK 412
 DB 23 IKLKEK 29
 |||||
 |||||
 RESULT 79
 AAY76347
 ID AAY76347 standard; protein; 45 AA.
 XX
 AC AAY76347;
 XX
 DT 23-MAR-2000 (first entry)
 XX
 DE Fragment of human secreted protein encoded by gene 48.
 XX
 KW Human; secreted protein; cancer; tumour; developmental abnormality;
 KW foetal deficiency; blood disorder; immune system disorder; inflammation;
 KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
 KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
 KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
 KW digestive disorder; endocrine disorder; infection; AIDS; leukaemia;
 KW therapy.
 XX
 KW Homo sapiens.
 OS
 XX WO9958660-A1.
 FN
 XX 18-NOV-1999.
 PD
 XX 06-MAY-1999; 99WO-US009847.
 PP
 XX 12-MAY-1998; 98US-0085093P.
 PR 12-MAY-1998; 98US-0085094P.
 PR 12-MAY-1998; 98US-0085105P.
 PR 12-MAY-1998; 98US-0085180P.
 PR 18-MAY-1998; 98US-0085906P.
 PR

PR 18-MAY-1998; 98US-0085920P.
 PR 18-MAY-1998; 98US-0085921P.
 PR 18-MAY-1998; 98US-0085922P.
 PR 18-MAY-1998; 98US-0085923P.
 PR 18-MAY-1998; 98US-0085924P.
 PR 18-MAY-1998; 98US-0085925P.
 PR 18-MAY-1998; 98US-0085926P.
 PR 18-MAY-1998; 98US-0085928P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Ruben SM, Florence K, Ni J, Rosen CA, Carter KC, Moore PA;
 PI Olsen HS, Shi Y, Young PE, Wei P, Brewer LA, Soppet DR, Lafleur DW;
 PI Endress GA, Ebner R;
 XX
 DR WPI; 2000-062296/05.
 XX
 XX New isolated human genes and the secreted polypeptides they encode,
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders.
 XX
 PS Disclosure; Page 452; 475pp; English.
 XX
 CC AAZ65250 to AAZ65350 represent 97 isolated human secreted protein genes.
 CC AAY76124 to AAY76223 are the secreted proteins encoded by the 97 human
 CC genes. This sequence represents a fragment of one of the human secreted
 CC proteins. The genes and their corresponding secreted polypeptides are
 CC useful for preventing, treating or ameliorating medical conditions, e.g.
 CC by protein or gene therapy. Also pathological conditions can be diagnosed
 CC by determining the amount of the new polypeptides in a sample or by
 CC determining the presence of mutations in the new genes. Specific uses are
 CC described for each of the 97 genes, based on which tissues they are most
 CC highly expressed in, and include developmental abnormalities and foetal
 CC treatment of cancer, tumours, developmental abnormalities and foetal
 CC deficiencies, blood disorders, diseases of the immune system, autoimmune
 CC diseases, inflammation, allergies, Alzheimer's and cognitive disorders,
 CC schizophrenia, arthritis, asthma, psoriasis, sepsis, skin disorders,
 CC atherosclerosis, diabetes, cardiovascular disorders, kidney disorders,
 CC digestive/endocrine disorders, infections and AIDS. The polypeptides are
 CC also useful for identifying their binding partners. The sequences shown
 CC in AAY76224 to AAY76424 represent fragments of the secreted proteins
 XX
 SQ Sequence 45 AA;
 Query Match 1.3%; Score 7; DB 3; Length 45;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 EGAQERP 18
 DB 34 EGAQERP 40
 |||||
 |||||
 RESULT 80
 AAB56241
 ID AAB56241 standard; protein; 45 AA.
 XX
 AC AAB56241;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Human secreted protein sequence encoded by gene 8 SEQ ID NO:335.
 XX
 KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; gene therapy; pathological condition;
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
 KW cerebrovascular disorder; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; skin aging; food additive; preservative.
 XX

OS Homo sapiens.
XX WO200070042-A1.
XX 23-NOV-2000.
XX PF 11-MAY-2000; 2000WO-US012788.
XX PR 13-MAY-1999; 99US-01340689.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
PI Duan RD, Florence KA, Soppet DR;
XX WPI; 2000-679828/66.
XX Isolated nucleic acid molecule encoding a human secreted protein is used
PT in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure; Page 1010; 1065pp; English.
XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the
CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: immunosuppressive;
CC antithrombotic; antineoplastic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
CC virucide; fungicide; and ophthalmological. The human secreted
CC polynucleotides and proteins can be used to prevent, treat or ameliorate
CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC dogs, chickens or sheep. They are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. Disorders which
CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC fungi and ocular disorders e.g. corneal infection. The proteins can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The proteins can also be used as a
CC food additive or preservative to increase or decrease storage
CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
CC in the exemplification of the present invention
XX
XX SQ Sequence 45 AA;
Query Match 1.3%; Score 7; DB 3; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 EGAQERP 18
DB 34 EGAQERP 40
RESULT 81
ADE11987
ID ADE11987 standard; protein; 45 AA.
XX
XX ADE11987;
XX 29-JAN-2004 (first entry)
XX Human secreted polypeptide #241.
XX
XX Secreted protein; cancer; liver disorder; hepatitis; neural disorder;
XX Alzheimer's disease; human.
XX Synthetic.
XX Homo sapiens.

XX US2003100051-A1.
XX 29-MAY-2003.
XX 10-SEP-2001; 2001US-00948783.
XX 12-MAY-1998; 98US-0085093P.
XX 12-MAY-1998; 98US-0085094P.
XX 12-MAY-1998; 98US-0085105P.
XX 12-MAY-1998; 98US-0085180P.
XX 12-MAY-1998; 98US-0085906P.
XX 18-MAY-1998; 98US-0085920P.
XX 18-MAY-1998; 98US-0085921P.
XX 18-MAY-1998; 98US-0085922P.
XX 18-MAY-1998; 98US-0085923P.
XX 18-MAY-1998; 98US-0085924P.
XX 18-MAY-1998; 98US-0085925P.
XX 18-MAY-1998; 98US-0085927P.
XX 18-MAY-1998; 98US-0085928P.
XX 06-MAY-1999; 99WO-US009847.
XX 10-NOV-1999; 99US-00437658.
XX 11-SEP-2000; 2000US-0231846P.
XX 23-JUN-2001; 2001US-00892877.
XX (RUBE/) RUBEN S M.
XX (FLO/) FLORENCE K A.
XX (NIJ/) NI J.
XX (ROSE/) ROSEN C A.
XX (CART/) CARTER K C.
XX (MOOR/) MOORE P A.
XX (OLSE/) OLSEN H S.
XX (SHI/) SHI Y.
XX (YOUN/) YOUNG P E.
XX (WEI/) WEI Y.
XX (BREW/) BREWER L A.
XX (SOPP/) Soppet D R.
XX (LAF/) LAFLEUR D W.
XX (ENDR/) ENDRESS G A.
XX (ERNE/) ERNER R.
XX (BIRS/) BIRSE C E.
XX Ruben SM, Florence KA, Ni J, Rosen CA, Carter KC, Moore PA;
PI Olsen HS, Shi Y, Young PE, Wei Y, Brewer LA, Soppet DR, Lafleur DW;
PI Endress GA, Ebner R, Birse CE;
XX WPI, 2003-801210/75.
XX New nucleic acid molecule, useful for preparing a medicament for
PT preventing, treating or ameliorating a medical condition e.g. Cancer,
PT liver disorders or neural disorders.
XX Claim 11; SEQ ID NO 359; 453pp; English.
XX The invention relates to human secreted polypeptides and the
CC polynucleotides encoding them. The sequences are useful for preparing
CC medicaments for preventing, treating or ameliorating medical conditions
CC e.g., cancer, liver disorders such as hepatitis or neural disorders such
CC as Alzheimer's disease. This sequence represents a human secreted
CC polypeptide of the invention.
XX
XX SQ Sequence 45 AA;
Query Match 1.3%; Score 7; DB 7; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 EGAQERP 18
DB 34 EGAQERP 40
RESULT 82

PN W09620953-A2.
XX 11-JUL-1996.
XX 22-DEC-1995; 95WO-US016854.
XX 30-DEC-1994; 94US-00366479.
XX (UTNC-) UNIV NORTH CAROLINA.
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX WPI; 1996-333940/33.
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX to produce and detect viral antibodies in a mammal and for immunisation.
XX Claim 1; Page 42; 50pp; English.
XX The sequences given in AAW03428-40 represent synthetic peptides with
XX hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
XX 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
XX HDV protein. Circular dichroism spectroscopy has demonstrated that this
XX protein has a strong concentration-dependent tendency for alpha-helical
XX coiled-coil complexes, with a Tm in excess of 80 deg. C. This peptide
XX demonstrates exceptionally strong and broadly reactive antigenic activity
XX and expresses conformational epitopes of HDV. It is thought to self-
XX assemble into a multimeric structure composed of four or more peptide
XX chains. The sequences given in W032429-40 represent analogs of delta12-
XX 60(Y). These peptides may be used to detect antibodies to HDV in a
XX mammal, and to produce neutralising antibodies to immunise animals
XX against HDV
XX
SQ Sequence 50 AA;
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 28 KKIKKLE 34
RESULT 84
AAW03429
ID AAW03429 standard; peptide; 50 AA.
XX AAW03429;
XX AC AAW03429;
XX 14-FEB-1997 (first entry)
XX DE HDV antigenic peptide analog, delta12-60(Y)/S22C.
XX KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX KW coiled-coil dimerisation domain; epitope; multimer.
XX OS Synthetic.
XX XX
XX W09620953-A2.
XX 11-JUL-1996.
XX 22-DEC-1995; 95WO-US016854.
XX 30-DEC-1994; 94US-00366479.
XX (UTNC-) UNIV NORTH CAROLINA.
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX WPI; 1996-333940/33.
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX to produce and detect viral antibodies in a mammal and for immunisation.
XX Claim 1; Page 42; 50pp; English.
XX The sequences given in AAW03428-40 represent synthetic peptides with
XX hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
XX 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
XX HDV protein. Circular dichroism spectroscopy has demonstrated that this
XX protein has a strong concentration-dependent tendency to for alpha-
XX helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
XX peptide demonstrates exceptionally strong and broadly reactive antigenic
XX activity and expresses conformational epitopes of HDV. It is thought to
XX self-assemble into a multimeric structure composed of four or more
XX peptide chains. The sequences given in W032429-40 represent analogs of
XX delta12-60(Y). These peptides may be used to detect antibodies to HDV in
XX a mammal, and to produce neutralising antibodies to immunise animals
XX against HDV
XX
SQ Sequence 50 AA;
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 28 KKIKKLE 34
RESULT 83
AAW03428
ID AAW03428 standard; peptide; 50 AA.
XX AAW03428;
XX 14-FEB-1997 (first entry)
XX DE Delta hepatitis virus derived peptide, delta12-60(Y).
XX KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX KW coiled-coil dimerisation domain; epitope; multimer.
XX OS Synthetic.
XX XX

AAW03431
ID AAW03431 standard; peptide; 50 AA.
XX AAW03431;
XX 14-FEB-1997 (first entry)
XX HDV antigenic peptide analog, delta12-60(Y)/Fr.
XX DE Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX KW coiled-coil dimerisation domain; epitope; multimer.
XX OS Synthetic.
XX W09620953-A2.
XX 11-JUL-1996.
XX 22-DEC-1995; 95WO-US016854.
XX 30-DEC-1994; 94US-00366479.
XX (UTNC-) UNIV NORTH CAROLINA.
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX WPI; 1996-333940/33.
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX to produce and detect viral antibodies in a mammal and for immunisation.
XX Claim 1; Page 34; 50pp; English.
XX The sequences given in AAW03428-40 represent synthetic peptides with
XX hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
XX 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
XX HDV protein. Circular dichroism spectroscopy has demonstrated that this
XX protein has a strong concentration-dependent tendency to for alpha-
XX helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
XX peptide demonstrates exceptionally strong and broadly reactive antigenic
XX activity and expresses conformational epitopes of HDV. It is thought to
XX self-assemble into a multimeric structure composed of four or more
XX peptide chains. The sequences given in W032429-40 represent analogs of
XX delta12-60(Y). These peptides may be used to detect antibodies to HDV in
XX a mammal, and to produce neutralising antibodies to immunise animals
XX against HDV
XX
SQ Sequence 50 AA;
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 28 KKIKKLE 34
RESULT 83
AAW03428
ID AAW03428 standard; peptide; 50 AA.
XX AAW03428;
XX 14-FEB-1997 (first entry)
XX DE Delta hepatitis virus derived peptide, delta12-60(Y).
XX KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX KW coiled-coil dimerisation domain; epitope; multimer.
XX OS Synthetic.
XX XX

PT Synthetic peptide with Delta hepatitis virus antigenic activity - useful
to produce and detect viral antibodies in a mammal and for immunisation.
XX
PS Claim 1; Page 33; 50pp; English.
XX
CC The sequences given in AAW03428-40 represent synthetic peptides with
hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
CC peptide demonstrates exceptionally strong and broadly reactive antigenic
CC activity and expresses conformational epitopes of HDV. It is thought to
CC self-assemble into a multimeric structure composed of four or more
CC peptide chains. The sequences given in W032429-40 represent analogs of
CC delta12-60(Y). These peptides may be used to detect antibodies to HDV in
CC a mammal, and to produce neutralising antibodies to immunise animals
CC against HDV
XX
SQ Sequence 50 AA;
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 28 KKIKKLE 34
RESULT 85
AAW03438
ID AAW03438 standard; peptide; 50 AA.
AC AAW03438;
XX
DT 14-FEB-1997 (first entry)
XX
DE HDV antigenic peptide analog, delta12-60(Y)/Fe.
XX
KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
KW coiled-coil dimerisation domain; epitope; multimer.
XX
OS Synthetic.
XX
PN W09620953-A2.
XX
PD 11-JUL-1996.
XX
PF 22-DEC-1995; 95WO-US016854.
XX
PR 30-DEC-1994; 94US-00366479.
XX
PA (UYNC-) UNIV NORTH CAROLINA.
XX
PI Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX
DR WPI; 1996-333940/33.
XX
PT Synthetic peptide with Delta hepatitis virus antigenic activity - useful
to produce and detect viral antibodies in a mammal and for immunisation.
XX
PS Claim 1; Page 37; 50pp; English.
XX
CC The sequences given in AAW03428-40 represent synthetic peptides with
hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
CC peptide demonstrates exceptionally strong and broadly reactive antigenic
CC activity and expresses conformational epitopes of HDV. It is thought to
CC self-assemble into a multimeric structure composed of four or more
CC peptide chains. The sequences given in W032429-40 represent analogs of
CC delta12-60(Y). These peptides may be used to detect antibodies to HDV in
CC a mammal, and to produce neutralising antibodies to immunise animals
CC against HDV
XX
SQ Sequence 50 AA;
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 28 KKIKKLE 34
RESULT 86
AAW03439
ID AAW03439 standard; peptide; 50 AA.
AC AAW03439;
XX
DT 14-FEB-1997 (first entry)
XX
DE HDV antigenic peptide analog, delta12-60(Y)/Ta.
XX
KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
KW coiled-coil dimerisation domain; epitope; multimer.
XX
OS Synthetic.
XX
PN W09620953-A2.
XX
PD 11-JUL-1996.
XX
PF 22-DEC-1995; 95WO-US016854.
XX
PR 30-DEC-1994; 94US-00366479.
XX
PA (UYNC-) UNIV NORTH CAROLINA.
XX
PI Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX
DR WPI; 1996-333940/33.
XX
PT Synthetic peptide with Delta hepatitis virus antigenic activity - useful
to produce and detect viral antibodies in a mammal and for immunisation.
XX
PS Claim 1; Page 38; 50pp; English.
XX
CC The sequences given in AAW03428-40 represent synthetic peptides with
hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
CC peptide demonstrates exceptionally strong and broadly reactive antigenic
CC activity and expresses conformational epitopes of HDV. It is thought to
CC self-assemble into a multimeric structure composed of four or more
CC peptide chains. The sequences given in W032429-40 represent analogs of
CC delta12-60(Y). These peptides may be used to detect antibodies to HDV in
CC a mammal, and to produce neutralising antibodies to immunise animals
CC against HDV
XX
SQ Sequence 50 AA;
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 28 KKIKKLE 34

RESULT 87
AAW03437
ID AAW03437 standard; peptide; 50 AA.

AC AAW03437;
XX
DT 14-FEB-1997 (first entry)
XX
DB HDV antigenic peptide analog, delta12-60(Y)/Na.
XX
KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
KW coiled-coil dimerisation domain; epitope; multimer.
XX
OS Synthetic.

XX
XX W09620953-A2.
XX
XX 11-JUL-1996.
XX
XX 22-DEC-1995; 95WO-US016854.
XX
XX 30-DEC-1994; 94US-00366479.
XX
XX (UYNC-) UNIV NORTH CAROLINA.

XX
PI Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX
XX WPI; 1996-333940/33.

XX
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX to produce and detect viral antibodies in a mammal and for immunisation.
XX
XX Claim 1; Page 37; 50pp; English.

XX
CC The sequences given in AAW03428-40 represent synthetic peptides with
CC hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
CC 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
CC peptide demonstrates exceptionally strong and broadly reactive antigenic
CC activity and expresses conformational epitopes of HDV. It is thought to
CC self-assemble into a multimeric structure composed of four or more
CC peptide chains. The sequences given in W032429-40 represent analogs of
CC delta12-60(Y). These peptides may be used to detect antibodies to HDV in
CC a mammal, and to produce neutralising antibodies to immunise animals
CC against HDV

XX
SQ Sequence 50 AA;

Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 404 KKIKKLE 410
Db 28 KKIKKLE 34

RESULT 88
AAW03434
ID AAW03434 standard; peptide; 50 AA.

AC AAW03434;
XX
XX 14-FEB-1997 (first entry)
XX
XX HDV antigenic peptide analog; delta12-60(Y)/Jal.

XX
KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
KW coiled-coil dimerisation domain; epitope; multimer.

XX
OS Synthetic.
XX
XX W09620953-A2.
XX
XX 11-JUL-1996.

XX
XX 22-DEC-1995; 95WO-US016854.
XX
XX 30-DEC-1994; 94US-00366479.
XX
XX (UYNC-) UNIV NORTH CAROLINA.

XX
PI Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX
XX WPI; 1996-333940/33.

XX
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX to produce and detect viral antibodies in a mammal and for immunisation.
XX
XX Claim 1; Page 35; 50pp; English.

XX
CC The sequences given in AAW03428-40 represent synthetic peptides with
CC hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
CC 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
CC peptide demonstrates exceptionally strong and broadly reactive antigenic
CC activity and expresses conformational epitopes of HDV. It is thought to
CC self-assemble into a multimeric structure composed of four or more
CC peptide chains. The sequences given in W032429-40 represent analogs of
CC delta12-60(Y). These peptides may be used to detect antibodies to HDV in
CC a mammal, and to produce neutralising antibodies to immunise animals
CC against HDV

XX
SQ Sequence 50 AA;

Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 404 KKIKKLE 410
Db 28 KKIKKLE 34

RESULT 89
AAW03436
ID AAW03436 standard; peptide; 50 AA.

XX
XX AAW03436;

XX
DT 14-FEB-1997 (first entry)
XX
DE HDV antigenic peptide analog, delta12-60(Y)/Le.

XX
KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
KW coiled-coil dimerisation domain; epitope; multimer.

XX
OS Synthetic.
XX
XX W09620953-A2.
XX
XX 11-JUL-1996.

XX
XX 22-DEC-1995; 95WO-US016854.
XX
XX 30-DEC-1994; 94US-00366479.
XX
XX (UYNC-) UNIV NORTH CAROLINA.

XX
PI Lemon SM, Erickson BW, Wang JG, Rozzelle J;

XX WPI; 1996-333940/33.
XX
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
PT to produce and detect viral antibodies in a mammal and for immunisation.
XX
XX Claim 1; Page 36; 50pp; English.
XX
XX The sequences given in AAW03428-40 represent synthetic peptides with
CC hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
CC 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
CC peptide demonstrates exceptionally strong and broadly reactive antigenic
CC activity and expresses conformational epitopes of HDV. It is thought to
CC self-assemble into a multimeric structure composed of four or more
CC peptide chains. The sequences given in W032429-40 represent analogs of
CC delta12-60(Y). These peptides may be used to detect antibodies to HDV in
CC a mammal, and to produce neutralising antibodies to immunise animals,
XX against HDV
XX
XX Sequence 50 AA;
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 28 KKIKKLE 34
RESULT 90
AAW03440
ID AAW03440 standard; peptide; 50 AA.
AC AAW03440;
DT 14-FEB-1997 (first entry)
XX HDV antigenic peptide analog, delta12-60(Y)-Cons.
XX Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX coiled-coil dimerisation domain; epitope; multimer.
XX Synthetic.
XX WO9620953-A2.
XX 11-JUL-1996.
XX 22-DEC-1995; 95WO-US016854.
XX 30-DEC-1994; 94US-00366479.
XX (UYNC-) UNIV NORTH CAROLINA.
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX WPI; 1996-333940/33.
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
PT to produce and detect viral antibodies in a mammal and for immunisation.
XX
XX Claim 1; Page 38; 50pp; English.
XX The sequences given in AAW03428-40 represent synthetic peptides with
CC hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
CC 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
XX peptide demonstrates exceptionally strong and broadly reactive antigenic
XX activity and expresses conformational epitopes of HDV. It is thought to
XX self-assemble into a multimeric structure composed of four or more
XX peptide chains. The sequences given in W032429-40 represent analogs of
XX delta12-60(Y). These peptides may be used to detect antibodies to HDV in
XX a mammal, and to produce neutralising antibodies to immunise animals,
XX against HDV
XX
XX Sequence 50 AA;
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 28 KKIKKLE 34
RESULT 91
ABG10316
ID ABG10316 standard; protein; 51 AA.
XX ABG10316;
AC ABG10316;
DT 13-FEB-2002 (first entry)
XX Novel human diagnostic protein #10307.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS74503.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 20; SEQ ID NO 40675; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 51 AA;

Query Match 1.3%; Score 7; DB 4; Length 51;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 247 AREEEK 253
 Db 14 AREEEK 20

RESULT 92
 ABG23757
 ID ABG23757 standard; protein; 52 AA.

XX AC ABG23757;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #23748.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PT 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX N-PSDB; AAS87944.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX PS Claim 20; SEQ ID NO 54116; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (II) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 52 AA;

Query Match 1.3%; Score 7; DB 4; Length 52;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 477 QGSLTDS 483
 Db 36 QGSLTDS 42

RESULT 93
 AAW03441
 ID AAW03441 standard; peptide; 57 AA.

XX AC AAW03441;

XX DT 16-OCT-2003 (revised)

XX DT 14-FEB-1997 (first entry)

XX DE HDV antigen fragment, residues 9-65.

XX KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
 KW coiled-coil dimerisation domain; epitope; multimer.

XX OS Hepatitis D virus.

XX PH Key Location/Qualifiers

XX PT Peptide 4..16

XX PT /label= Segment A
 PT /note= "Corresponds to HDag residues 12-24"

XX PT Peptide 17..41

XX PT /label= Segment B
 PT /note= "Corresponds to HDag residues 25-49"

XX PT Peptide 42..52

XX PT /label= Segment C
 PT /note= "Corresponds to HDag residues 50-60"

XX PN WO9620953-A2.

XX PD 11-JUL-1996.

XX PF 22-DEC-1995; 95WO-US016854.

XX PR 30-DEC-1994; 94US-00366479.

XX PA (UYN-) UNIV NORTH CAROLINA.

XX PI Lemon SM, Erickson BW, Wang JG, Rozzelle J;

XX WPI; 1996-333940/33.

XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
 PT to produce and detect viral antibodies in a mammal and for immunisation.

XX Example 10; Page 29; 50pp; English.

XX This sequence represents residues 9-65 of the hepatitis delta virus (HDV)
 CC antigen (HDag) sequence. This sequence was used in the design of the
 CC sequences given in AAW03441-47. These sequences represent synthetic
 CC peptides with hepatitis delta virus (HDV) antigenic activity. They were
 CC used to raise murine monoclonal antibodies which may be used to immunise
 CC animals against HDV, and to detect anti-HDV antibodies. (Updated on 16-
 CC OCT-2003 to standardise OS field)

XX SQ Sequence 57 AA;

Query Match 1.3%; Score 7; DB 2; Length 57;
 Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 404 KKIKKLE 410
Db 31 KKIKKLE 37

RESULT 94
ABG27967
ID ABG27967 standard; protein; 57 AA.
XX AC ABG27967;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #27958.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX XW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX PA Drmanac RT, Liu C, Tang YT;
XX PI WPI; 2001-639362/73.
XX DR N-PSDB; AAS92154.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 58325; 103pp; English.

Query Match 1.3%; Score 7; DB 4; Length 57;
Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy 404 QGSLTDS 483
Db 40 QGSLTDS 46

RESULT 95
ADC00699
ID ADC00699 standard; protein; 60 AA.
XX AC ADC00699;
XX DT 04-DEC-2003 (first entry)
XX DE Enterohaemorrhagic E. coli O157:H7-specific protein SEQ ID NO: 744.
XX KW enterohaemorrhagic; anti-bacterial.
XX OS Escherichia coli; O157:H7.
XX XW JP2002355074-A.
XX PN 10-DEC-2002.
XX PD 24-JAN-2002; 2002JP-00015959.
XX PF 24-JAN-2001; 2001JP-00112010.
XX PR (UVTS-) UNIV TSUKUBA.
XX PA WPI; 2003-451640/43.
XX PT Enterohaemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule
PT and a polypeptide and its use, a polypeptide, a vector and a host cell.
XX Claim 3; SEQ ID NO 744; 2067pp; Japanese.
XX The invention relates to a novel enterohaemorrhagic Escherichia coli
CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
CC has anti-bacterial activity. The polypeptide can be used in detection
CC and/or treatment of O157:H7 infection. The nucleotide sequence of the
CC genome of Enterohaemorrhagic E coli O157:H7 was determined. The present
CC sequence represents an E. coli O157:H7-specific polypeptide of the
CC invention.

Query Match 1.3%; Score 7; DB 7; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy 494 AQAPSSP 500
Db 48 AQAPSSP 54

RESULT 96
ABG02022
ID ABG02022 standard; protein; 70 AA.
XX AC ABG02022;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #2013.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX XW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PR 11-OCT-2001.

Query Match 1.3%; Score 7; DB 4; Length 57;
Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

```

XX 30-MAR-2001; 2001WO-US008631.
XX
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS66209.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 20; SEQ ID NO 32381; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel biodiversity
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 70 AA;
XX
Query Match 1.3%; Score 7; DB 4; Length 70;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 477 QGSLTDS 483
DB 53 QGSLTDS 59
|||||
RESULT 97
ID ABG27972
AC ABG27972 standard; protein; 72 AA.
XX
XX ABG27972;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #27963.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX

```

```

XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS92159.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 20; SEQ ID NO 58331; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 72 AA;
XX
Query Match 1.3%; Score 7; DB 4; Length 72;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 477 QGSLTDS 483
DB 55 QGSLTDS 61
|||||
RESULT 98
ID AAO01466
AC AAO01466 standard; protein; 73 AA.
XX
XX AAO01466;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 15358.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US004927.
XX

```

```

XX 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX DR N-PSDB; AAI81397.
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX Claim 20; SEQ ID NO 15358; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93941) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: the sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 73 AA;
Query Match 1.3%; Score 7; DB 4; Length 73;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AQERPSQ 20
Db 58 AQERPSQ 64
|||||

RESULT 99
ABG26987
ID ABG26987 standard; protein; 75 AA.
XX AC ABG26987;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #26978.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX WO200175067-A2.
XX PN 11-OCT-2001.
XX PD 30-MAR-2001; 2001WO-US008631.
XX PF 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS91174.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT responsible for genetic disorders or other traits and to assess

diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX Claim 20; SEQ ID NO 57346; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 75 AA;
Query Match 1.3%; Score 7; DB 4; Length 75;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 QGSLTDS 483
Db 66 QGSLTDS 72
|||||

RESULT 100
ABG28171
ID ABG28171 standard; protein; 75 AA.
XX AC ABG28171;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #28162.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX WO200175067-A2.
XX PN 11-OCT-2001.
XX PD 30-MAR-2001; 2001WO-US008631.
XX PF 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS92358.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess

```

PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 58530; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG00017 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 75 AA;

Query Match 1.3%; Score 7; DB 4; Length 75;
 Best Local Similarity 100.0%; Pred.No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 477 QGSLTDS 483
 DB 58 QGSLTDS 64
 |||||

Search completed: June 7, 2004, 14:45:20
 Job time : 59 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:44:26 ; Search time 20 seconds
(without alignments)
1368.088 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 530
Sequence: 1 KSPXQPEAGPEGAQERPSQ.....APSTEASQGTGQPTPSARA 530

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 300 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCITUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	530	100.0	530	4	US-08-979-608A-8
2	530	100.0	530	4	US-09-517-849-8
3	530	100.0	530	4	US-09-616-289-8
4	429	80.9	546	4	US-09-616-289-44
5	172	32.5	557	4	US-08-979-608A-5
6	172	32.5	557	4	US-09-517-849-5
7	172	32.5	557	4	US-09-616-289-5
8	19	3.6	386	2	US-08-968-751-6
9	8	1.5	165	4	US-09-252-991A-19490
10	8	1.5	260	2	US-08-879-561-7
11	7	1.3	7	4	US-08-979-608A-41
12	7	1.3	7	4	US-09-517-849-41
13	7	1.3	7	4	US-09-616-289-41
14	7	1.3	77	3	US-09-446-350-25
15	7	1.3	105	4	US-09-732-210-434
16	7	1.3	124	4	US-09-489-039A-10493
17	7	1.3	127	4	US-08-849-303-19
18	7	1.3	142	4	US-09-489-039A-9418
19	7	1.3	159	3	US-09-146-950-20
20	7	1.3	172	4	US-08-079-030-117
21	7	1.3	187	4	US-09-328-352-7281
22	7	1.3	197	3	US-09-146-950-18
23	7	1.3	211	4	US-09-134-001C-5175
24	7	1.3	261	2	US-08-879-561-1
25	7	1.3	267	4	US-09-489-039A-13457
26	7	1.3	274	1	US-08-287-959-3
27	7	1.3	277	4	US-09-857-498A-7

28	7	1.3	281	4	US-09-134-000C-3980	Sequence 3980, Ap
29	7	1.3	294	4	US-09-489-039A-8327	Sequence 8327, Ap
30	7	1.3	302	4	US-09-328-352-4508	Sequence 4508, Ap
31	7	1.3	311	4	US-09-252-991A-28792	Sequence 28792, A
32	7	1.3	312	1	US-08-285-440-4	Sequence 4, Appli
33	7	1.3	312	1	US-08-630-349-4	Sequence 4, Appli
34	7	1.3	314	4	US-09-252-991A-18599	Sequence 18599, A
35	7	1.3	330	4	US-09-252-991A-28531	Sequence 28531, A
36	7	1.3	341	3	US-09-008-465-1	Sequence 1, Appli
37	7	1.3	341	4	US-09-528-959-1	Sequence 1, Appli
38	7	1.3	346	4	US-09-252-991A-27167	Sequence 27167, A
39	7	1.3	364	4	US-09-252-991A-27209	Sequence 27209, A
40	7	1.3	388	4	US-09-252-991A-21572	Sequence 21572, A
41	7	1.3	389	4	US-09-543-681A-7318	Sequence 7318, Ap
42	7	1.3	411	4	US-09-252-991A-24632	Sequence 24632, A
43	7	1.3	422	4	US-09-252-991A-21571	Sequence 21571, A
44	7	1.3	446	4	US-09-252-991A-25318	Sequence 25318, A
45	7	1.3	473	4	US-09-252-991A-16904	Sequence 16904, A
46	7	1.3	474	4	US-09-543-681A-6752	Sequence 6752, Ap
47	7	1.3	495	3	US-08-962-859A-2	Sequence 2, Appli
48	7	1.3	526	4	US-09-252-991A-20533	Sequence 20533, A
49	7	1.3	532	1	US-08-285-440-5	Sequence 5, Appli
50	7	1.3	532	1	US-08-630-349-5	Sequence 5, Appli
51	7	1.3	557	4	US-09-138-277C-3	Sequence 3, Appli
52	7	1.3	558	1	US-08-285-440-6	Sequence 6, Appli
53	7	1.3	581	1	US-08-630-349-6	Sequence 6, Appli
54	7	1.3	581	4	US-09-499-522-18	Sequence 18, Appli
55	7	1.3	582	4	US-09-269-939A-12	Sequence 12, Appli
56	7	1.3	582	4	US-09-252-991A-27626	Sequence 27626, A
57	7	1.3	630	4	US-09-499-522-16	Sequence 16, Appli
58	7	1.3	630	4	US-09-269-939A-10	Sequence 10, Appli
59	7	1.3	649	4	US-09-499-522-14	Sequence 14, Appli
60	7	1.3	649	4	US-09-269-939A-8	Sequence 8, Appli
61	7	1.3	712	4	US-09-543-681A-5392	Sequence 5392, Ap
62	7	1.3	1024	4	US-09-562-737-47	Sequence 47, Appli
63	7	1.3	1075	4	US-09-252-991A-18387	Sequence 18387, A
64	7	1.3	1141	1	US-08-131-365B-54	Sequence 54, Appli
65	7	1.3	1141	2	US-08-668-123-54	Sequence 54, Appli
66	7	1.3	1530	4	US-09-976-594-736	Sequence 736, App
67	7	1.3	1857	1	US-08-287-959-1	Sequence 1, Appli
68	7	1.3	3340	4	US-09-252-991A-23568	Sequence 23568, A
69	6	1.1	6	2	US-08-310-912A-138	Sequence 138, App
70	6	1.1	6	3	US-09-301-085-138	Sequence 138, App
71	6	1.1	6	5	PCT-US95-04589-138	Sequence 138, App
72	6	1.1	7	4	US-09-989-789-1988	Sequence 1988, Ap
73	6	1.1	9	1	US-08-339-152A-1	Sequence 1, Appli
74	6	1.1	9	2	US-08-007-999B-1	Sequence 1, Appli
75	6	1.1	9	2	US-08-689-276A-1	Sequence 1, Appli
76	6	1.1	10	3	US-08-159-339A-830	Sequence 830, App
77	6	1.1	11	3	US-08-836-075A-109	Sequence 109, App
78	6	1.1	12	3	US-08-747-599A-16	Sequence 16, Appli
79	6	1.1	15	4	US-08-979-608A-29	Sequence 29, Appli
80	6	1.1	15	4	US-09-517-849-29	Sequence 29, Appli
81	6	1.1	15	4	US-09-616-289-29	Sequence 29, Appli
82	6	1.1	16	1	US-08-408-120-15	Sequence 15, Appli
83	6	1.1	17	4	US-09-459-958-2	Sequence 2, Appli
84	6	1.1	18	6	5304631-6	Patent No. 5304631
85	6	1.1	19	2	US-08-802-322-10	Sequence 10, Appli
86	6	1.1	19	4	US-09-832-161-10	Sequence 10, Appli
87	6	1.1	20	4	US-09-834-759-544	Sequence 544, App
88	6	1.1	21	4	US-09-851-843A-113	Sequence 543, App
89	6	1.1	23	3	US-08-854-050-113	Sequence 113, App
90	6	1.1	23	3	US-08-854-050-113	Sequence 113, App
91	6	1.1	23	4	US-09-430-323-113	Sequence 113, App
92	6	1.1	23	4	US-08-912-951-113	Sequence 114, App
93	6	1.1	23	4	US-09-402-181B-233	Sequence 233, App
94	6	1.1	23	4	US-09-721-456-233	Sequence 233, App
95	6	1.1	25	2	US-08-507-598-4	Sequence 4, Appli
96	6	1.1	25	2	US-08-507-750-4	Sequence 4, Appli
97	6	1.1	25	3	US-08-764-532A-4	Sequence 4, Appli
98	6	1.1	25	3	US-08-764-538-4	Sequence 4, Appli
99	6	1.1	25	3	US-08-872-859-4	Sequence 4, Appli
100	6	1.1	25	3	US-08-872-859-4	Sequence 4, Appli

101	6	1.1	30	3	US-08-927-219-41	Sequence 41, Appl	174	6	1.1	96	3	US-08-654-482-11	Sequence 11, Appl
102	6	1.1	32	2	US-08-708-620A-3	Sequence 3, Appl	175	6	1.1	98	2	US-08-341-843B-13	Sequence 13, Appl
103	6	1.1	36	1	US-08-441-534A-7	Sequence 7, Appl	176	6	1.1	98	2	US-08-427-497E-18	Sequence 18, Appl
104	6	1.1	36	1	US-08-441-534A-8	Sequence 8, Appl	177	6	1.1	100	1	US-08-473-981A-11	Sequence 11, Appl
105	6	1.1	36	1	US-08-629-752-7	Sequence 7, Appl	178	6	1.1	100	2	US-08-474-087-11	Sequence 11, Appl
106	6	1.1	36	1	US-08-629-752-8	Sequence 8, Appl	179	6	1.1	100	3	US-09-034-916-13	Sequence 13, Appl
107	6	1.1	36	2	US-08-802-991-7	Sequence 7, Appl	180	6	1.1	105	2	US-08-690-011A-15	Sequence 15, Appl
108	6	1.1	36	2	US-08-802-991-8	Sequence 8, Appl	181	6	1.1	105	4	US-09-299-495F-15	Sequence 15, Appl
109	6	1.1	36	4	US-09-079-030-61	Sequence 61, Appl	182	6	1.1	105	4	US-09-107-532A-5344	Sequence 5344, Ap
110	6	1.1	36	4	US-08-488-161-14	Sequence 14, Appl	183	6	1.1	106	4	US-09-069-023-10	Sequence 10, Appl
111	6	1.1	38	2	US-08-726-306A-101	Sequence 101, App	184	6	1.1	106	4	US-09-621-976-4439	Sequence 4439, Ap
112	6	1.1	38	3	US-09-273-685-14	Sequence 14, Appl	185	6	1.1	107	4	US-09-765-298A-2	Sequence 2, Appl
113	6	1.1	38	5	PCT-US95-11934-14	Sequence 14, Appl	186	6	1.1	109	4	US-09-574-141A-72	Sequence 72, Appl
114	6	1.1	40	1	US-08-129-129-6	Sequence 6, Appl	187	6	1.1	109	4	US-09-325-932A-48	Sequence 48, Appl
115	6	1.1	41	1	US-08-453-924-8	Sequence 8, Appl	188	6	1.1	110	4	US-09-134-000C-6399	Sequence 6399, Ap
116	6	1.1	41	1	US-08-208-008C-11	Sequence 11, Appl	189	6	1.1	112	4	US-09-107-532A-4572	Sequence 4572, Ap
117	6	1.1	45	1	US-08-339-152A-25	Sequence 25, Appl	190	6	1.1	114	2	US-08-822-262-3	Sequence 3, Appl
118	6	1.1	45	2	US-08-007-939B-8	Sequence 8, Appl	191	6	1.1	114	4	US-08-679-493A-187	Sequence 187, App
119	6	1.1	46	2	US-08-689-276A-8	Sequence 8, Appl	192	6	1.1	114	4	US-09-252-991A-18514	Sequence 18514, A
120	6	1.1	46	2	US-09-205-258-583	Sequence 583, App	193	6	1.1	115	4	US-09-376-330-22	Sequence 22, Appl
121	6	1.1	51	4	US-09-621-976-7200	Sequence 7200, Ap	194	6	1.1	117	4	US-09-134-000C-5275	Sequence 5275, Ap
122	6	1.1	51	2	US-08-621-976-7200	Sequence 7200, Ap	195	6	1.1	117	4	US-09-134-000C-5589	Sequence 5589, Ap
123	6	1.1	57	2	US-08-164-292B-33	Sequence 33, Appl	196	6	1.1	117	4	US-09-134-000C-5589	Sequence 5589, Ap
124	6	1.1	57	3	US-08-845-623-33	Sequence 33, Appl	197	6	1.1	118	4	US-09-252-991A-23679	Sequence 23679, A
125	6	1.1	57	3	US-08-815-927-33	Sequence 33, Appl	198	6	1.1	122	3	US-08-956-107B-19	Sequence 19, Appl
126	6	1.1	57	4	US-09-103-330-33	Sequence 33, Appl	198	6	1.1	124	4	US-09-134-000C-6744	Sequence 6744, Ap
127	6	1.1	57	4	US-08-435-242-33	Sequence 33, Appl	199	6	1.1	125	4	US-09-621-976-4822	Sequence 4822, Ap
128	6	1.1	60	4	US-09-621-976-5357	Sequence 5357, Ap	200	6	1.1	128	2	US-08-928-926A-3	Sequence 3, Appl
129	6	1.1	65	4	US-09-328-352-7594	Sequence 7594, Ap	201	6	1.1	128	3	US-09-212-149-3	Sequence 3, Appl
130	6	1.1	66	3	US-09-370-473-8	Sequence 8, Appl	202	6	1.1	129	4	US-09-732-210-1307	Sequence 1307, Ap
131	6	1.1	67	4	US-09-621-976-7214	Sequence 7214, Ap	203	6	1.1	130	4	US-09-732-210-1308	Sequence 1308, Ap
132	6	1.1	68	4	US-09-621-976-5028	Sequence 7215, Ap	204	6	1.1	132	4	US-09-716-129-57	Sequence 57, Appl
133	6	1.1	73	4	US-08-580-043B-1	Sequence 5028, Ap	205	6	1.1	133	4	US-09-732-210-82	Sequence 82, Appl
134	6	1.1	74	3	US-08-851-843A-187	Sequence 2, Appl	206	6	1.1	133	4	US-09-489-039A-13255	Sequence 13255, A
135	6	1.1	74	3	US-08-974-549A-306	Sequence 187, App	207	6	1.1	134	4	US-09-252-991A-24342	Sequence 24342, A
136	6	1.1	74	3	US-08-836-075A-10	Sequence 306, App	208	6	1.1	135	4	US-09-328-352-5695	Sequence 5695, Ap
137	6	1.1	74	3	US-08-854-050-187	Sequence 10, Appl	209	6	1.1	137	4	US-09-732-210-122	Sequence 122, App
138	6	1.1	74	3	US-09-430-323-187	Sequence 187, App	210	6	1.1	137	6	5202428-9	Patent No. 5202428
139	6	1.1	74	4	US-08-635-886C-198	Sequence 187, App	211	6	1.1	138	3	US-08-836-075A-60	Sequence 60, Appl
140	6	1.1	74	4	US-09-402-181B-306	Sequence 198, App	212	6	1.1	140	4	US-09-540-236-2611	Sequence 2611, Ap
141	6	1.1	74	4	US-08-974-690C-198	Sequence 306, App	213	6	1.1	141	3	US-09-230-944-4	Sequence 4, Appl
142	6	1.1	74	4	US-09-721-456-306	Sequence 198, App	214	6	1.1	141	4	US-09-873-233A-4	Sequence 4, Appl
143	6	1.1	75	1	US-07-934-017-1	Sequence 306, App	215	6	1.1	142	4	US-09-543-681A-4478	Sequence 4478, Ap
144	6	1.1	75	1	US-09-252-991A-29599	Sequence 1, Appl	216	6	1.1	142	4	US-09-543-681A-6036	Sequence 6036, Ap
145	6	1.1	75	4	US-09-732-210-833	Sequence 1, Appl	217	6	1.1	142	4	US-09-489-039A-8382	Sequence 8382, Ap
146	6	1.1	76	1	US-08-168-941-1	Sequence 833, App	218	6	1.1	142	4	US-09-489-039A-11181	Sequence 11181, A
147	6	1.1	79	4	US-09-134-000C-5581	Sequence 1, Appl	219	6	1.1	143	4	US-09-387-418A-17	Sequence 17, Appl
148	6	1.1	80	4	US-09-107-532A-5588	Sequence 5581, Ap	220	6	1.1	144	4	US-09-732-210-1406	Sequence 1406, Ap
149	6	1.1	82	2	US-08-690-011A-13	Sequence 5588, Ap	221	6	1.1	144	4	US-09-489-039A-11477	Sequence 11477, A
150	6	1.1	82	3	US-08-905-223-398	Sequence 13, Appl	222	6	1.1	145	4	US-09-489-039A-10614	Sequence 10614, A
151	6	1.1	82	4	US-09-299-495F-13	Sequence 398, App	223	6	1.1	146	2	US-07-885-0895-37	Sequence 37, Appl
152	6	1.1	83	4	US-09-732-210-1584	Sequence 13, Appl	224	6	1.1	146	4	US-09-252-991A-21410	Sequence 21410, A
153	6	1.1	83	4	US-09-732-210-1587	Sequence 1584, Ap	225	6	1.1	147	4	US-09-252-991A-21269	Sequence 21269, A
154	6	1.1	83	4	US-09-134-000C-5875	Sequence 1587, Ap	226	6	1.1	148	4	US-09-621-976-4708	Sequence 4708, Ap
155	6	1.1	83	4	US-09-621-976-7069	Sequence 5875, Ap	227	6	1.1	149	4	US-09-489-039A-10664	Sequence 10664, A
156	6	1.1	84	1	US-08-510-039-2	Sequence 7069, Ap	228	6	1.1	151	4	US-09-489-039A-13318	Sequence 13318, A
157	6	1.1	84	1	US-09-748-510-2	Sequence 2, Appl	229	6	1.1	152	1	US-08-046-583-11	Sequence 11, Appl
158	6	1.1	88	4	US-09-134-000C-5567	Sequence 2, Appl	230	6	1.1	152	1	US-08-261-677-11	Sequence 11, Appl
159	6	1.1	89	4	US-08-618-485B-2	Sequence 5567, Ap	231	6	1.1	152	1	US-08-384-556A-7	Sequence 7, Appl
160	6	1.1	90	4	US-09-107-532A-6712	Sequence 2, Appl	232	6	1.1	152	1	US-08-384-556A-7	Sequence 7, Appl
161	6	1.1	91	4	US-09-732-210-1218	Sequence 6712, Ap	233	6	1.1	152	1	US-08-592-936B-17	Sequence 17, Appl
162	6	1.1	92	4	US-09-489-039A-11823	Sequence 1218, Ap	234	6	1.1	152	2	US-07-885-0895-36	Sequence 36, Appl
163	6	1.1	93	4	US-08-936-165A-381	Sequence 11823, A	235	6	1.1	152	2	US-08-331-355A-11	Sequence 11, Appl
164	6	1.1	94	2	US-08-737-716-10	Sequence 381, App	236	6	1.1	152	2	US-08-331-355A-23	Sequence 23, Appl
165	6	1.1	94	4	US-09-099-041A-31	Sequence 10, Appl	237	6	1.1	152	2	US-09-111-573-17	Sequence 17, Appl
166	6	1.1	94	4	US-09-245-281-31	Sequence 31, Appl	238	6	1.1	152	4	US-09-071-035-128	Sequence 128, App
167	6	1.1	94	4	US-08-618-485B-3	Sequence 31, Appl	239	6	1.1	152	5	PCT-US94-12364-11	Sequence 11, Appl
168	6	1.1	94	4	US-09-207-359B-31	Sequence 3, Appl	240	6	1.1	152	5	PCT-US94-12364-23	Sequence 23, Appl
169	6	1.1	94	4	US-09-340-620A-31	Sequence 31, Appl	241	6	1.1	152	5	PCT-US95-07753-2	Sequence 2, Appl
170	6	1.1	94	4	US-09-252-991A-26448	Sequence 31, Appl	242	6	1.1	152	5	PCT-US95-07753-7	Sequence 7, Appl
171	6	1.1	94	4	US-09-865-364-31	Sequence 26448, A	243	6	1.1	153	4	US-09-199-6377A-213	Sequence 213, App
172	6	1.1	94	4	US-09-673-395A-415	Sequence 31, Appl	244	6	1.1	154	4	US-09-252-991A-20709	Sequence 20709, A
173	6	1.1	96	3	US-09-258-373-22	Sequence 415, App	245	6	1.1	154	4	US-09-107-532A-5411	Sequence 5411, Ap
						Sequence 22, Appl	246	6	1.1	155	1	US-08-328-961-8	Sequence 8, Appl


```

247 1 1.1 155 1 US-08-462-397-8
248 3 1.1 155 3 US-08-989-251-39
249 3 1.1 155 3 US-09-340-250-39
250 6 1.1 155 4 US-09-528-108-39
251 6 1.1 155 4 US-09-732-210-1237
252 6 1.1 157 4 US-09-252-991A-20847
253 6 1.1 160 4 US-09-252-991A-28043
254 6 1.1 160 4 US-09-198-452A-1244
255 6 1.1 162 4 US-09-198-452A-1190
256 6 1.1 164 4 US-09-621-976-5067
257 6 1.1 165 4 US-09-252-991A-27342
258 6 1.1 166 4 US-09-134-000C-5854
259 6 1.1 168 4 US-09-252-991A-30271
260 6 1.1 168 4 US-09-107-532A-3745
261 6 1.1 169 3 US-08-928-941D-35
262 6 1.1 169 4 US-09-280-590A-45
263 6 1.1 169 4 US-09-892-398-45
264 6 1.1 172 4 US-09-107-532A-5616
265 6 1.1 174 4 US-09-328-352-7878
266 6 1.1 175 4 US-09-543-681A-8208
267 6 1.1 175 4 US-09-543-681A-8328
268 6 1.1 176 4 US-09-387-418A-16
269 6 1.1 176 4 US-09-252-991A-21594
270 6 1.1 179 4 US-09-252-991A-24507
271 6 1.1 180 4 US-09-252-991A-18339
272 6 1.1 180 4 US-09-252-991A-23183
273 6 1.1 180 4 US-09-252-991A-26449
274 6 1.1 181 1 US-08-034-245-14
275 6 1.1 181 4 US-09-325-932A-180
276 6 1.1 181 4 US-09-252-991A-19502
277 6 1.1 181 4 US-09-134-000C-5492
278 6 1.1 182 4 US-09-198-452A-164
279 6 1.1 182 4 US-09-543-681A-5698
280 6 1.1 182 4 US-09-134-000C-6084
281 6 1.1 183 4 US-09-071-035-126
282 6 1.1 183 4 US-09-489-039A-10037
283 6 1.1 183 4 US-09-134-000C-4534
284 6 1.1 185 4 US-09-198-452A-1209
285 6 1.1 185 4 US-09-418-710-64
286 6 1.1 187 4 US-09-252-991A-29442
287 6 1.1 188 4 US-09-107-532A-7007
288 6 1.1 191 3 US-08-989-251-41
289 6 1.1 191 3 US-09-340-250-41
290 6 1.1 191 4 US-09-528-108-41
291 6 1.1 192 4 US-09-489-039A-10990
292 6 1.1 192 4 US-09-134-000C-3560
293 6 1.1 194 4 US-09-252-991A-22009
294 6 1.1 195 4 US-09-252-991A-21451
295 6 1.1 195 4 US-09-489-039A-7491
296 6 1.1 196 4 US-09-540-236-3127
297 6 1.1 199 4 US-09-252-991A-18507
298 6 1.1 202 4 US-08-311-731A-13
299 6 1.1 204 4 US-09-230-670C-8
300 6 1.1 207 4 US-09-252-991A-28701

```

ALIGNMENTS

```

RESULT 1
US-08-979-608A-8
; Sequence 8, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
;               Lees, Robert S.
;               Law, Simon W.
;               Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
;                   BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
;                   TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESS: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-Nov. 6355451-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-979-608A-8

Query Match      100.0%; Score 530; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KSSPGQPEAGPEGAQPSQAAPAVEAGPGSSQAPKPEGAQARTAGSGALRVSELS 60
Db 1 KSSPGQPEAGPEGAQPSQAAPAVEAGPGSSQAPKPEGAQARTAGSGALRVSELS 60
OY 61 RQLSDILSTYCVNNQGGEDGAGPEPAEDAEKSTYVARNGEPEPTPVYGEKPS 120
Db 61 RQLSDILSTYCVNNQGGEDGAGPEPAEDAEKSTYVARNGEPEPTPVYGEKPS 120
OY 121 KGDPTNTEIRQSDVGDHRRPOEKKAQGLGKEITLLMOTLNTLSTPEKLAALCKY 180
Db 121 KGDPTNTEIRQSDVGDHRRPOEKKAQGLGKEITLLMOTLNTLSTPEKLAALCKY 180
OY 181 AELLEHHNSOKMKLLQKQKQSLVQEKDHLRGHSHKAVLARSKLESICRELQHNRLK 240
Db 181 AELLEHHNSOKMKLLQKQKQSLVQEKDHLRGHSHKAVLARSKLESICRELQHNRLK 240
OY 241 EGVQRAPEEERKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENMELASRLKKLIQY 300
Db 241 EGVQRAPEEERKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENMELASRLKKLIQY 300
OY 301 ELRBEHDVFKHNDLQOOLVDKLOQAQEMKAEERHOREKDFLLKAEVESQRMCELM 360
Db 301 ELRBEHDVFKHNDLQOOLVDKLOQAQEMKAEERHOREKDFLLKAEVESQRMCELM 360
OY 361 KQOETHLQKQALVTEKPEEPQNTLSKSSSEVFTTFKQEMENVTKKIKKLETTMYRSRW 420
Db 361 KQOETHLQKQALVTEKPEEPQNTLSKSSSEVFTTFKQEMENVTKKIKKLETTMYRSRW 420
OY 421 ESSNKALLEAEAEKTVRDKELEGQVKIQRLEKLCRALQTERNDLNKGVQDLSAGGQSL 480
Db 421 ESSNKALLEAEAEKTVRDKELEGQVKIQRLEKLCRALQTERNDLNKGVQDLSAGGQSL 480
OY 481 TDSGPRERPEPGQAQPSPPVTEAPCPVGPAPSTEASGQTPGPTPSARA 530
Db 481 TDSGPRERPEPGQAQPSPPVTEAPCPVGPAPSTEASGQTPGPTPSARA 530

```

Db 481 TDSGPERPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 530

RESULT 2

US-09-517-849-8

; Sequence 8, Application US/09517849

; Patent No. 5605588

; GENERAL INFORMATION:

; APPLICANT: Lees, Ann M.

; Lees, Robert S.

; Law, Simon W.

; Arjona, Anibal A.

; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN

; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND

; TREATING ATHEROSCLEROSIS

; NUMBER OF SEQUENCES: 42

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/517,849

; FILING DATE: 02-Mar-2000

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/379,608

; FILING DATE: 26-NOV-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Myers, Louis

; REGISTRATION NUMBER: 35,965

; REFERENCE/DOCKET NUMBER: 10797-003001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/542-5070

; TELEFAX: 617/542-8906

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 530 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-09-517-849-8

Query Match 100.0%; Score 530; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 KSSPGQPEAGPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 530
DB 1 KSSPGQPEAGPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 530
QY 61 RQLEDILSTYCVDDNNGGPGEDGAQGEPAEPEDAERKSRITYVARNGEPPTPVVYGEKPS 120
DB 61 RQLEDILSTYCVDDNNGGPGEDGAQGEPAEPEDAERKSRITYVARNGEPPTPVVYGEKPS 120
QY 121 KGPDPTEIRQSDVGDHRRPQEKKAAGLKEITLLMTLTNTLSTPEKLAALCKKY 180
DB 121 KGPDPTEIRQSDVGDHRRPQEKKAAGLKEITLLMTLTNTLSTPEKLAALCKKY 180
QY 181 AELLEHRNSQOMKLLQKQSQVLQVKDHLRGHSHKAVLARSKLSLCELOQRNRSK 240
DB 181 AELLEHRNSQOMKLLQKQSQVLQVKDHLRGHSHKAVLARSKLSLCELOQRNRSK 240
QY 241 EGVQARAREEERKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300
DB 241 EGVQARAREEERKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300
```

```
QY 301 ELREEHIDKVPKHDLQQOLVDKILQQAQEMLKBAERHOREKDFLLKEAVSORMCELM 360
DB 301 ELREEHIDKVPKHDLQQOLVDKILQQAQEMLKBAERHOREKDFLLKEAVSORMCELM 360
QY 361 KQOETHLKOALALYTEKFEFPQNTLSKSSEVFTTTPKQEMKMTKKIKLEKETTMYRSW 420
DB 361 KQOETHLKOALALYTEKFEFPQNTLSKSSEVFTTTPKQEMKMTKKIKLEKETTMYRSW 420
QY 421 ESSNKALLEWABEKTVRDKLEGLQVKIORLKLGRALQTERNDLNKRYQDLSAGQGS 480
DB 421 ESSNKALLEWABEKTVRDKLEGLQVKIORLKLGRALQTERNDLNKRYQDLSAGQGS 480
QY 481 TDSGPERPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 530
DB 481 TDSGPERPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 530
```

RESULT 3

US-09-616-289-8

; Sequence 8, Application US/09616289

; Patent No. 5632923

; GENERAL INFORMATION:

; APPLICANT: Lees, Ann M.

; APPLICANT: Lees, Robert S.

; APPLICANT: Law, Simon W.

; APPLICANT: Arjona, Anibal A.

; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING

; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING

; TITLE OF INVENTION: ATHEROSCLEROSIS

; FILE REFERENCE: 10797-004001

; CURRENT APPLICATION NUMBER: US/09/616,289

; CURRENT FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: US 09/517,849

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: US 08/379,608

; PRIOR FILING DATE: 1997-11-26

; PRIOR APPLICATION NUMBER: US 60/031,930

; PRIOR FILING DATE: 1996-11-27

; PRIOR APPLICATION NUMBER: US 60/048,547

; PRIOR FILING DATE: 1997-06-03

; NUMBER OF SEQ ID NOS: 53

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8

; LENGTH: 530

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-616-289-8

Query Match 100.0%; Score 530; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 KSSPGQPEAGPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 530
DB 1 KSSPGQPEAGPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 530
QY 61 RQLEDILSTYCVDDNNGGPGEDGAQGEPAEPEDAERKSRITYVARNGEPPTPVVYGEKPS 120
DB 61 RQLEDILSTYCVDDNNGGPGEDGAQGEPAEPEDAERKSRITYVARNGEPPTPVVYGEKPS 120
QY 121 KGPDPTEIRQSDVGDHRRPQEKKAAGLKEITLLMTLTNTLSTPEKLAALCKKY 180
DB 121 KGPDPTEIRQSDVGDHRRPQEKKAAGLKEITLLMTLTNTLSTPEKLAALCKKY 180
QY 181 AELLEHRNSQOMKLLQKQSQVLQVKDHLRGHSHKAVLARSKLSLCELOQRNRSK 240
DB 181 AELLEHRNSQOMKLLQKQSQVLQVKDHLRGHSHKAVLARSKLSLCELOQRNRSK 240
QY 241 EGVQARAREEERKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300
DB 241 EGVQARAREEERKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300
```

QY 301 ELREEHIDKVFHKDLQOQVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
DB 301 ELREEHIDKVFHKDLQOQVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
QY 361 KOQETHLKOQALYTEKPEFQNTLSKSSSEVFTTFKQEMEKTKYIKKLEKETTMYRSRW 420
DB 361 KOQETHLKOQALYTEKPEFQNTLSKSSSEVFTTFKQEMEKTKYIKKLEKETTMYRSRW 420
QY 421 ESSNKALLMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLS 480
DB 421 ESSNKALLMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLS 480
QY 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 530
DB 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 530

RESULT 4

US-09-616-289-44
; Sequence 44, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-616-289-44

Query Match 80.9%; Score 429; DB 4; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPQAGPQAPPAVEAGPGSSQAPRKGPAQARTAGSGLRDVSEELS 60
DB 17 KSSPGQPEAGPQAGPQAPPAVEAGPGSSQAPRKGPAQARTAGSGLRDVSEELS 76
QY 61 RQLEDILSTYCVNNQGGEDGAGQEPEDAEKSRITYVARNGEPPTPVVGEKPS 120
DB 77 RQLEDILSTYCVNNQGGEDGAGQEPEDAEKSRITYVARNGEPPTPVVGEKPS 136
QY 121 KGPNTETIRQSDVGDHRDHPQEKKAQKGLQKBITLLMQTLNTLSTPEEKLAALCKY 180
DB 137 KGPNTETIRQSDVGDHRDHPQEKKAQKGLQKBITLLMQTLNTLSTPEEKLAALCKY 196
QY 181 AELLEHRNSQKMLLQKQSQOLVQKDLHGEHSKAVLARSKLESCLREIQRNRSIK 240
DB 197 AELLEHRNSQKMLLQKQSQOLVQKDLHGEHSKAVLARSKLESCLREIQRNRSIK 256
QY 241 EECVQARSEEEKREKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKRLIKIYQ 300
DB 257 EECVQARSEEEKREKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKRLIKIYQ 316
QY 301 ELREEHIDKVFHKDLQOQVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360

DB 317 ELREEHIDKVFHKDLQOQVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376
QY 361 KOQETHLKOQALYTEKPEFQNTLSKSSSEVFTTFKQEMEKTKYIKKLEKETTMYRSRW 420
DB 377 KOQETHLKOQALYTEKPEFQNTLSKSSSEVFTTFKQEMEKTKYIKKLEKETTMYRSRW 436
QY 421 ESSNKALLMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLS 480
DB 437 ESSNKALLMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLS 496
QY 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 530
DB 497 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 546

RESULT 5

US-08-979-608A-5
; Sequence 5, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-08-979-608A-5

Query Match 32.5%; Score 172; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 1.3e-149;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 264 TLNDIQLQMEQHNRNSKLRQENMELAEKRLIKIYQYELREEHIDKVFHKDLQOQVDA 323
DB 280 TLNDIQLQMEQHNRNSKLRQENMELAEKRLIKIYQYELREEHIDKVFHKDLQOQVDA 339

Qy 324 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKOQALALYTEKFEFQON 383
Db 340 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKOQALALYTEKFEFQON 399
Qy 384 TLSKSSEVFTTFKQEMEKNTKKIKLEKETTMYRSWESSNKALLEMAEKT 435
Db 400 TLSKSSEVFTTFKQEMEKNTKKIKLEKETTMYRSWESSNKALLEMAEKT 451

RESULT 6
US-09-517-849-5
; Sequence 5, Application US/09517849
; Patent No. 6605588
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; MEDIUM TYPE: diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/517,849
; FILING DATE: 02-Mar-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608
; FILING DATE: 26-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-517-849-5

Query Match 32.5%; Score 172; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 1.3e-149;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 264 TLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQYELREHIDKVFKHDLQOQLVDA 323
Db 280 TLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQYELREHIDKVFKHDLQOQLVDA 339
Qy 324 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKOQALALYTEKFEFQON 383
Db 340 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKOQALALYTEKFEFQON 399
Qy 384 TLSKSSEVFTTFKQEMEKNTKKIKLEKETTMYRSWESSNKALLEMAEKT 435
Db 400 TLSKSSEVFTTFKQEMEKNTKKIKLEKETTMYRSWESSNKALLEMAEKT 451

RESULT 8
US-08-968-751-6
; Sequence 6, Application US/08968751
; Patent No. 5948643
; GENERAL INFORMATION:
; APPLICANT: Rubinfeld, Bonnie
; APPLICANT: Polakis, Paul G.
; APPLICANT: Ligenfelter, Carol
; APPLICANT: Vuong, Terilyn T.
; TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONYX Pharmaceuticals, Inc.
; STREET: 3031 Research Drive
; CITY: Richmond
; STATE: CA
; COUNTRY: USA
; ZIP: 94806
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

RESULT 7
US-09-616-289-5
; Sequence 5, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-09-616-289-5

Query Match 32.5%; Score 172; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 1.3e-149;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 264 TLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQYELREHIDKVFKHDLQOQLVDA 323
Db 280 TLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQYELREHIDKVFKHDLQOQLVDA 339
Qy 324 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKOQALALYTEKFEFQON 383
Db 340 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKOQALALYTEKFEFQON 399
Qy 384 TLSKSSEVFTTFKQEMEKNTKKIKLEKETTMYRSWESSNKALLEMAEKT 435
Db 400 TLSKSSEVFTTFKQEMEKNTKKIKLEKETTMYRSWESSNKALLEMAEKT 451

RESULT 8
US-08-968-751-6
; Sequence 6, Application US/08968751
; Patent No. 5948643
; GENERAL INFORMATION:
; APPLICANT: Rubinfeld, Bonnie
; APPLICANT: Polakis, Paul G.
; APPLICANT: Ligenfelter, Carol
; APPLICANT: Vuong, Terilyn T.
; TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONYX Pharmaceuticals, Inc.
; STREET: 3031 Research Drive
; CITY: Richmond
; STATE: CA
; COUNTRY: USA
; ZIP: 94806
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/968,751
/ FILING DATE:
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Giotta, Gregory
/ REGISTRATION NUMBER: 32,028
/ REFERENCE/DOCKET NUMBER: ONYX1024 GG
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (510) 262-8710
/ TELEFAX: (510) 222-9758
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 386 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-968-751-6

Query Match 3.6%; Score 19; DB 2; Length 386;
Best Local Similarity 100.0%; Pred. No. 2.9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 LNTLSTPEKLAALCKKYA 181
Db 135 LNTLSTPEKLAALCKKYA 153

RESULT 9
US-09-252-991A-19490
/ Sequence 19490, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 19490
/ LENGTH: 165
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19490

Query Match 1.5%; Score 8; DB 4; Length 155;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 DRDHRPQ 144
Db 40 DRDHRPQ 47

RESULT 10
US-08-879-561-7
/ Sequence 7, Application US/08879561
/ Patent No. 5817482
/ GENERAL INFORMATION:
/ APPLICANT: Bandman, Olga
/ APPLICANT: Hillman, Jennifer L.
/ APPLICANT: Hawkins, Phillip R.
/ APPLICANT: Guegler, Karl J.
/ APPLICANT: Corley, Neil C.
/ TITLE OF INVENTION: DISEASE RELATED NUCLEOTIDE KINASES
/ NUMBER OF SEQUENCES: 12
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Incyte Pharmaceuticals, Inc.

/ STREET: 3174 Porter Drive
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/879,561
/ FILING DATE: Herewith
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Billings, Lucy J.
/ REGISTRATION NUMBER: 36,749
/ REFERENCE/DOCKET NUMBER: PF-0325 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415-855-0555
/ TELEFAX: 415-845-4166
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 260 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: GenBank
/ CLONE: 1480198
/ US-08-879-561-7

Query Match 1.5%; Score 8; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 QRAREEEE 252
Db 188 QRAREEEE 195

RESULT 11
US-08-979-608A-41
/ Sequence 41, Application US/08979608A
/ Patent No. 6355451
/ GENERAL INFORMATION:
/ APPLICANT: Lees, Ann M.
/ Lees, Robert S.
/ Law, Simon W.
/ Arjona, Anibal A.
/ TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
/ BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
/ TREATING ATHEROSCLEROSIS
/ NUMBER OF SEQUENCES: 42
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Fish & Richardson P.C.
/ STREET: 225 Franklin Street
/ CITY: Boston
/ STATE: MA
/ COUNTRY: USA
/ ZIP: 02110-2804
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/979,608A
/ FILING DATE: 26-No. 6355451-1997

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/048,547
FILING DATE: 03-JUN-1997
APPLICATION NUMBER: US 60/031,930
FILING DATE: 27-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-08-979-608A-41

Query Match
Best Local Similarity 1.3%; Score 7; DB 4; Length 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 12
US-09-517-849-41
Sequence 41, Application US/09517849
Patent No. 6605588
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fast-SEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/517,849
FILING DATE: 02-Mar-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/979,608
FILING DATE: 26-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-517-849-41

Query Match
Best Local Similarity 1.3%; Score 7; DB 4; Length 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 13
US-09-616-289-41
Sequence 41, Application US/09616289
Patent No. 6632923
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
APPLICANT: Lees, Robert S.
APPLICANT: Law, Simon W.
APPLICANT: Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
ATHEROSCLEROSIS
TITLE OF INVENTION: ATHEROSCLEROSIS
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/09/616,289
CURRENT FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 09/517,849
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 08/979,608
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 41
LENGTH: 7
TYPE: PRT
ORGANISM: Homo sapiens
US-09-616-289-41

Query Match
Best Local Similarity 1.3%; Score 7; DB 4; Length 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 14
US-09-146-950-25
Sequence 25, Application US/09146950A
Patent No. 6287808
GENERAL INFORMATION:
APPLICANT: Busfield, Samantha J.
TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
FAMILY AND USES THEREOF
FILE REFERENCE: 09404/057001
CURRENT APPLICATION NUMBER: US/09/146,950A
CURRENT FILING DATE: 1998-09-03
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 25
LENGTH: 77
TYPE: PRT
ORGANISM: Homo sapiens
US-09-146-950-25

Query Match
Best Local Similarity 1.3%; Score 7; DB 3; Length 77;

Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 145 EKXKAG 151
Db 71 EKXKAG 77

RESULT 15
US-09-732-210-434
; Sequence 434, Application US/09732210
; Patent No. 6573361
; GENERAL INFORMATION:
; APPLICANT: Bunkers, Greg J.
; APPLICANT: Lians, Jihong
; APPLICANT: Mittanck, Cindy A.
; APPLICANT: Seale, Jeffrey W.
; APPLICANT: Wu, Yomie S.
; TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
; FILE REFERENCE: 38-21(15036)B
; CURRENT APPLICATION NUMBER: US/09/732,210
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,513
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: US 60/169,340
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 1753
; SEQ ID NO 434
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Odontella sinensis
US-09-732-210-434

Query Match 1.3%; Score 7; DB 4; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LGKEITL 158
Db 30 LGKEITL 36

RESULT 16
US-09-489-039A-10493
; Sequence 10493, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10493
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10493

Query Match 1.3%; Score 7; DB 4; Length 124;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 LLKEAVE 352
Db 111 LLKEAVE 117

RESULT 17
US-08-849-303-19

; Sequence 19, Application US/08849303
; Patent No. 6680424
; GENERAL INFORMATION:
; APPLICANT: Ackinson, Howard J.
; APPLICANT: McPherson, Michael J.
; APPLICANT: Urwin, Peter E.
; TITLE OF INVENTION: MODIFIED PROTEINASE INHIBITORS
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/849,303
; FILING DATE: 21-MAY-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1321-1-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 127 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-849-303-19

Query Match 1.3%; Score 7; DB 4; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 BEGVORA 247
Db 27 BEGVORA 33

RESULT 18
US-09-489-039A-9418
; Sequence 9418, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9418
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9418

Query Match 1.3%; Score 7; DB 4; Length 142;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 157 TLLMQTL 163
Db 60 TLLMQTL 66

RESULT 19
US-09-146-950-20
; Sequence 20, Application US/09146950A
; Patent No. 6287808
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; FILE REFERENCE: 09404/057001
; CURRENT APPLICATION NUMBER: US/09/146,950A
; CURRENT FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-146-950-20

Query Match 1.3%; Score 7; DB 3; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 145 EKKKAGK 151
Db 153 EKKKAGK 159

RESULT 20
US-09-079-030-117
; Sequence 117, Application US/09079030
; Patent No. 6635623
; GENERAL INFORMATION:
; APPLICANT: Guevera, Jr., Juan G.
; APPLICANT: Hoogveen, Ron C.
; APPLICANT: Moore, Paul J.
; TITLE OF INVENTION: LIPOPROTEINS AS NUCLEIC ACID DELIVERY
; NUMBER OF SEQUENCES: 229
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,030
; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McMillian, Nabeela R.
; REGISTRATION NUMBER: P-43,363
; REFERENCE/DOCKET NUMBER: ARAG:003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 117:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 172 amino acids
; TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear
US-09-079-030-117

Query Match 1.3%; Score 7; DB 4; Length 172;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 281 KLRQKNM 287
Db 100 KLRQKNM 106

RESULT 21
US-09-328-352-7281
; Sequence 7281, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7281
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7281

Query Match 1.3%; Score 7; DB 4; Length 187;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 439 KELEGLO 445
Db 83 KELEGLO 89

RESULT 22
US-09-146-950-18
; Sequence 18, Application US/09146950A
; Patent No. 6287808
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; FILE REFERENCE: 09404/057001
; CURRENT APPLICATION NUMBER: US/09/146,950A
; CURRENT FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-146-950-18

Query Match 1.3%; Score 7; DB 3; Length 197;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 145 EKKKAGK 151
Db 191 EKKKAGK 197

RESULT 23
US-09-134-001C-5175
; Sequence 5175, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5175
LENGTH: 211
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5175

Query Match 1.3%; Score 7; DB 4; Length 211;
Best Local Similarity 100.0%; Pred. No. 1.9e+02; Indels 0;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 404 KKKKLE 410
Db 19 KKKKLE 25

RESULT 24
US-08-879-561-1
Sequence 1, Application US/08879561
Patent No. 5817482
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hillman, Jennifer L.
APPLICANT: Hawkins, Phillip R.
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: DISEASE RELATED NUCLEOTIDE KINASES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/879,561
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: Pf-0325 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: FIBROT01

APPLICANT: GARY BRETON et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.20040001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 13457
LENGTH: 267
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13457

Query Match 1.3%; Score 7; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 2.3e+02; Indels 0;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 493 GAQAPSS 499
Db 113 GAQAPSS 119

RESULT 26
US-08-287-959-3
Sequence 3, Application US/08287959
Patent No. 5639651
GENERAL INFORMATION:
APPLICANT: Weissbach, Lawrence
APPLICANT: Bernards, Andre
APPLICANT: Settleman, Jeffrey
TITLE OF INVENTION: GAP-RELATED GENE
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/287,959
FILING DATE: August 9, 1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul C.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/181001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070

CLONE: 56821
US-08-879-561-1

Query Match 1.3%; Score 7; DB 2; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.2e+02; Indels 0;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 245 QRAREEE 251
Db 189 QRAREEE 195

RESULT 25
US-09-489-039A-13457
Sequence 13457, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.20040001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 13457
LENGTH: 267
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13457

Query Match 1.3%; Score 7; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 2.3e+02; Indels 0;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 493 GAQAPSS 499
Db 113 GAQAPSS 119

RESULT 26
US-08-287-959-3
Sequence 3, Application US/08287959
Patent No. 5639651
GENERAL INFORMATION:
APPLICANT: Weissbach, Lawrence
APPLICANT: Bernards, Andre
APPLICANT: Settleman, Jeffrey
TITLE OF INVENTION: GAP-RELATED GENE
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/287,959
FILING DATE: August 9, 1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul C.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/181001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 274 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-287-959-3

Query Match 1.3%; Score 7; DB 1; Length 274;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 471 DLSAGGQ 477
Db 213 DLSAGGQ 219

RESULT 27
US-09-857-498A-7
Sequence 7, Application US/09857498A
Patent No. 6582947
GENERAL INFORMATION:
APPLICANT: Karlsson, Anna
APPLICANT: Johansson, Magnus
TITLE OF INVENTION: New medical use
FILE REFERENCE: multibacillus deoxyguanosine kinase
CURRENT APPLICATION NUMBER: US/09/857,498A
CURRENT FILING DATE: 2002-01-17
PRIOR APPLICATION NUMBER: SE 9804298-9
PRIOR FILING DATE: 1998-12-11
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 277
TYPE: PRT
ORGANISM: Homo sapiens
US-09-857-498A-7

Query Match 1.3%; Score 7; DB 4; Length 277;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 245 QRAREEE 251
Db 205 QRAREEE 211

RESULT 28
US-09-134-000C-3980
Sequence 3980, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3980
LENGTH: 281
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-134-000C-3980

Query Match 1.3%; Score 7; DB 4; Length 281;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 KLESCLR 230
Db 50 KLESCLR 56

RESULT 29
US-09-489-039A-8327
Sequence 8327, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2904001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 8327
LENGTH: 294
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8327

Query Match 1.3%; Score 7; DB 4; Length 294;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 ELLEHR 188
Db 19 ELLEHR 25

RESULT 30
US-09-328-352-4508
Sequence 4508, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4508
LENGTH: 302
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-4508

Query Match 1.3%; Score 7; DB 4; Length 302;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 345 FLKEAV 351
Db 102 FLKEAV 108

RESULT 31
US-09-252-991A-28792
Sequence 28792, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18

```

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28792
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28792

Query Match 1.3%; Score 7; DB 4; Length 311;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 QSDVEGD 137
DB 148 QSDVEGD 154

RESULT 32
US-08-285-440-4
; Sequence 4, Application US/08285440
; Patent No. 5532337
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/285,440
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 312 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEetical:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:

; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-4

Query Match 1.3%; Score 7; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 REBEKR 254
DB 115 REBEKR 121

RESULT 33
US-08-630-349-4
; Sequence 4, Application US/08630349
; Patent No. 5739008
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,349
; FILING DATE: April 10, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/285,440
; FILING DATE: August 4, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 312 amino acids
```

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-630-349-4

Query Match 1.3%; Score 7; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 REEERK 254
DB 115 REEERK 121

RESULT 34

US-09-252-991A-18599
; Sequence 18599, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18599
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18599

Query Match 1.3%; Score 7; DB 4; Length 314;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 QSDVEVD 137
DB 147 QSDVEVD 153

RESULT 35

US-09-252-991A-28531
; Sequence 28531, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28531
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28531

Query Match 1.3%; Score 7; DB 4; Length 330;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 RDHRRPQ 144

DB 259 RDHRRPQ 265

RESULT 36

US-09-008-465-1
; Sequence 1, Application US/09008465
; Patent No. 6174702
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN PINCH PROTEIN HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,465
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0460 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 341 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SEMVNOT04
; CLONE: 3540806
US-09-008-465-1

Query Match 1.3%; Score 7; DB 3; Length 341;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 KAKGLGK 154
DB 131 KAKGLGK 137

RESULT 37

US-09-528-959-1
; Sequence 1, Application US/09528959
; Patent No. 6379904
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN PINCH PROTEIN HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA: US/09/528,959
FILING DATE: 20-Mar-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/008,465
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0460 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 341 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SEMWNOT04
CLONE: 3540806
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-528-959-1

Query Match 1.3%; Score 7; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 148 KAKGLGK 154
Db 131 KAKGLGK 137
RESULT 38
US-09-252-991A-27167
; Sequence 27167, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27167
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27167

Query Match 1.3%; Score 7; DB 4; Length 346;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 AAPAVEA 27
Db 62 AAPAVEA 68
RESULT 39
US-09-252-991A-27209
; Sequence 27209, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27209
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27209

Query Match 1.3%; Score 7; DB 4; Length 388;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 487 RRPECPG 493
Db 76 RRPECPG 82
RESULT 41
US-09-543-681A-7318
; Sequence 7318, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; FILE REFERENCE: 2709.1002-001

RESULT 39
US-09-252-991A-27209
; Sequence 27209, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27209
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27209
Query Match 1.3%; Score 7; DB 4; Length 364;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 136 GDRDHR 142
Db 332 GDRDHR 338

RESULT 40
US-09-252-991A-21572
; Sequence 21572, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21572
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21572

Query Match 1.3%; Score 7; DB 4; Length 388;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 487 RRPECPG 493
Db 76 RRPECPG 82
RESULT 41
US-09-543-681A-7318
; Sequence 7318, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7318
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7318

Query Match 1.3%; Score 7; DB 4; Length 389;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 PTPVVG 115
Db 329 PTPVVG 335
|||||

RESULT 42
US-09-252-991A-24632
; Sequence 24632, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24632
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24632

Query Match 1.3%; Score 7; DB 4; Length 411;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PQQPEAG 10
Db 149 PQQPEAG 155
|||||

RESULT 43
US-09-252-991A-21571
; Sequence 21571, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21571
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21571

Query Match 1.3%; Score 7; DB 4; Length 411;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PQQPEAG 10
Db 149 PQQPEAG 155
|||||

RESULT 43
US-09-252-991A-21571
; Sequence 21571, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21571
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21571

Query Match 1.3%; Score 7; DB 4; Length 422;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 RRPEPG 493
Db 16 RRPEPG 22
|||||

RESULT 44
US-09-252-991A-25318
; Sequence 25318, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25318
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25318

Query Match 1.3%; Score 7; DB 4; Length 446;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 MELAERL 293
Db 415 MELAERL 421
|||||

RESULT 45
US-09-252-991A-16904
; Sequence 16904, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16904
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16904

Query Match 1.3%; Score 7; DB 4; Length 473;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GAQARTA 47
Db 418 GAQARTA 424
|||||

RESULT 46
US-09-543-681A-6752


```
SQ SEQUENCE 900 AA; 99136 MW; 1865EB38ACF9E615 CRC64;
Query Match 9.9%; Score 266.5; DB 3; Length 900;
Best Local Similarity 23.2%; Pred. No. 4.4e-05;
Matches 141; Conservative 103; Mismatches 240; Indels 123; Gaps 27;

QY 1 KSSPGQFAGPEGCAQ-----EPSP---QAAPAVEAGGSSQAPKPEGA--QARTASGA 51
DQ 101 QASPRQMDAPRSTASIERPFIADKASPNVSPRASADTPKXSQDTTQEPVPAFLDS 160
QY 52 LRDVSEELSRQLEIDILSTYCVNNQGGPGEGGAGQEPAPEDAKSRITYVARNGEPPTP 111
DQ 161 ADDIKEYEVAQSVSE-----KADSRPSTD--QPSPEITDEAPT-----TESKLEAPA 206
QY 112 VYGEKEPSKDPNTEERIQSDVGDGDRHRPQPKKAGKIGKEITILMOTL----- 163
DQ 207 EILAKTEPKSVBTEKARPDQGMKSDDTTPQNDQEIYAYVERIDALEAKLOYLAREAS 266
QY 164 -----NTLSTP-----EKKLAALCKKYAELLBEHRN--SOKQ-----MKLQKQKQSLVQVE 207
DQ 267 AARKEALSAPGSAEKLAEDQDQIQALQMEGKGNLASNEQKLTILKNLRKKQAEDEKD 326
QY 208 KDHURGHSHKAVLARSKLSLRC---ELQRHRSKKEEGVQVQARAESEKRVVTSHPQVT 264
DQ 327 MGNLKAAREKADREIENLRKEARHADELEKSNBLQKRLDQSQRELNYLRPEVKS-K-DTI 385
QY 265 LNDIQLOMEQHNER-----NSKLQENMELAEKLLKLEQYELREEHIDKVFHKDL 316
DQ 386 IAEIRSIQKATQADQVMSAKANDKAREQQ--RRITAEI-----EESVEALKIEKNL 435
QY 317 QOQLVDKALQQAQEMKAEERHQRKDF-----LKAEAVESOR----- 355
DQ 436 ---MADRKAQADELRKEAEKASEKAKALELEKAEVHMESKLEAMRTAEABASSGVTG 492
QY 356 ---MCELMKQETHLKKQALALYTFEPEFQNTLSKSVFTTFQEMKMTKKIKLEKET 413
DQ 493 DSQAKLLRQVET--LOSQYSIASENWQIETL--RSRIVNLEKERD-----ALQRES 542
QY 414 TMYRSRVSS-----NKALLEMAEKTVRDKELGELQVKIQRLEKLCR-----ALQTERND 464
DQ 543 DMRRKAREALARRNEEELEAKTLPNQEDVESYESQLDSLKRAEAEAEALAEARAD 602
QY 465 LNKRVQDLSAGGQSFLDSGPERPBGGAQAPSPRVTRAPCVGAP--STASGQTGP 522
DQ 603 FEKQKQAWEA--EKELIKERERDLSQG-----NRPR--SWLEGIPGGFFLKNEGSGPSP 655
QY 523 QEPTSA 529
DQ 656 QLSTAQR 662

RESULT 28
Q9UPV0 PRELIMINARY; PRT; 1455 AA.
AC Q9UPV0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE Hypothetical protein KIAA1052.
GN KIAA1052.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX [1]
RN TISSUE=Brain;
RC TISSUE=Brain;
RA MEDLINE=9937452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K., Hirose M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
```

```
RL DNA Res. 6:197-205 (1999).
DR EXBL; AB028975; BAA83004.1; -.
DR InterPro; IPR001202; WW_Reps_WWP.
DR Pfam; PF00397; WW; 1.
DR SMART; SM00456; WW; 1.
DR PROSITE; PS00020; WW_DOMAIN_2; 1.
RW Hypothetical protein.
SQ SEQUENCE 1455 AA; 163543 MW; 7F48093100C34819 CRC64;

Query Match 9.9%; Score 266.5; DB 4; Length 1455;
Best Local Similarity 21.6%; Pred. No. 7.1e-05;
Matches 132; Conservative 105; Mismatches 206; Indels 167; Gaps 20;

QY 7 PEAGPSGAQERSQAAPAVEAGCP-----GSSQAPKPEGAQARTASGALRDVSELSR 61
DQ 330 PKADPTG-----SEPAKASEKEAPETVDAGEGSRREAAKEPKKASALESGSDASQ 384
QY 62 QLE-----DILSTYCVNNQGGPGEGGAGQEPAPEDAKSRITYVARNGEPPTP 91
DQ 385 ELEISEHMKPEQLSDSIADSPKSFHGLDFGFTSRISEHLLDVLDFVLGACRQACQPL 444
QY 92 -----EDASKSRITYA-----RNGSP-----EPTPVVYGEKPSKDPNT 126
DQ 445 GIEDKDDSSODDELQSKQSKGLEERYHRLSPPLPHEERAQSPRSLATEEPPOGPEQ 504
QY 127 EETROSDEVDG-----RHRHREPOEK 146
DQ 505 PEWKEAEELGHSAAASLSLQSLQREQAPSPPAACEKGKHSQAEELGFGQGEAEDPEE 564
QY 147 KKAAGLGKKEITILMOTLNTLSTPE-----KLAALCKKYAELLBEHRNRSOKMKLQKQS 202
DQ 565 KVAVSPTPVPSPEVSTEPVADPEQLSEAAKXAMEAVAVLE-----QQRHLLSKOE 619
QY 203 QLVQEKDHLRGHSHKAVLA--RSKLESCLRELRHRSKKEGVQVQAREEEK-----RK 255
DQ 620 KMQOLREKLCQEEBEELRHQOKEQSLSLRERLQKALEEEA--RNRRESORLSWLRA 678
QY 256 EYTSHPQVTLNLIQLOME-----QNRNRSKLQENMELAEKLLKLEQYELR 303
DQ 679 QVQSSTQADEQIRAEQAEASLQKLEELSESQKAEASLEQKNQRMLEQLKHEIRASEKS 738
QY 304 EKHIDKVFHKDLQ---QLVDAKLOQAQEMLE-----AEERHOREKDFLKE 349
DQ 739 EQAALNAKKEKALQQLREQLGERKEAVATLSEKHSALERLCSSLEAKHREVSVSLQKK 798
QY 350 AVESQRMCELMKQ---QETHLKKQALALYTFEPEFQNTL--SKSSEVFTTTFQEMKMT 403
DQ 799 IQEAQKKEAQLKQKLCIGQVEHRVHQSYPVAGYEHLSLLREKQVEGEHERLDKM- 857
QY 404 KXIKKLEKTTMYRGRWESSNKALLEMAEKTVRDKELGELQVKIQRLEKLC-RALQTER 462
DQ 858 -----KEEHQOVMAKAREQVE-----AERKQRAELLGHLTGELERLQRAHERLETVR 906
QY 463 NDLNKRQVDL 472
DQ 907 QEQHKELEDL 916

RESULT 29
Q9TV62 PRELIMINARY; PRT; 1937 AA.
AC Q9TV62;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Myosin heavy chain 2b.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Landrace; TISSUE=Skeletal muscle;
```

```

RA Chikuni K., Tanabe R., Muroya S., Nakajima I.;
RT "Differences in molecular structure among the porcine myosin heavy
RL chain-2a, -2x, and -2b isoforms.";
RL Meat Sci., 57:311-317(2001);
DR EMBL; AB025261; BAA82145.1; --
DR HSSP; P13538; 2MYS.
DR GO; GO:0016459; C:myosin; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
SQ SEQUENCE 1937 AA; 223235 MW; BCC114C6824E0426 CRC64;

Query Match 9.7%; Score 262; DB 6; Length 1937;
Best Local Similarity 21.9%; Pred. No. 0.00015;
Matches 118; Conservative 78; Mismatches 172; Indels 170; Gaps 15;

Qy 25 VEAEPCSSAPKPECAQARTAGSALRDVSELSQLEDILSTYCVNNOGPGEDGA 84
Db 1096 IDEQALAMQLKKLEQART-----BELLEEIE----- 1125
Qy 85 QGPAEPEDAEKRTYVARNGEPEPTPVVYGEKPSKGDNTBEIROSDEVGDRHRRPQ 144
Db 1126 -AFSPAPAKQKQSDLSRELE-----ISERLEAGGATSAQI-----EMNKKREA 1171
Qy 145 EKKAKGLGKEITLLMOTLNTLSTPEKLAALCKY-----AELEHNSQKQKMLLOKK 200
Db 1172 EFCQREDELEATL-----QHEATAAALRKCHADSVAEELGEQIDNLRVKQKLEKE 1222
Qy 201 QSOLVQEKDLRGHSHKAVLARSKLESICREL-----QRNRSIKKEGVQAR- 248
Db 1223 KSELKGEIDDLASNETVSKAGNLEKMCRTLEDQLESEVTKKEEHQRLNELSAQKRL 1282
Qy 249 -----EEERKKEVTSH-FQVTLNDI 268
Db 1283 QTESGEFSRQDKEALVSLSRGKAFTQOIIEELKRLKEETKAKSALAHAVQSSRHDC 1342
Qy 269 QLQWEOHNER-----NSKLR-----QENMELAERLKGLIEQVELR 303
Db 1343 DLUREQYEEGEQAKAEIQRAMSKANSEVAQWRTKYETDAIQTREEEAKKLAQRLQDA 1402
Qy 304 EEHIDKVPKHKDLOQQQVDAKLAQAQAEMLKEAEERHOREKDFLKAEAVESQRMCELMKQ 363
Db 1403 EEHV-----EAVNAKAS-----LEKTKQLQNEVEDLMDVRSNAACAALDKK 1447
Qy 364 ETHLKQOALYTEKPEEPQNTLSKSEVFTFPKQEMKMTKKLKEKETMTWRSWESS 423
Db 1448 QNPKDILAEKMKHYEETQ-----AELEASQKESRSLSSTELFKVKNAYEES 1493
Qy 424 NKALLEAAEEKTVRDKLEGLQVK-----IORLEKLCRALQTERNDLNKRVDLSA 474
Db 1494 LDQLETLKRENKNLQQHISDLTQEIABGGKHIELEKVKQIQEKKSELQAALAEBA 1551

RESULT 30
Q9DGM4 PRELIMINARY; PRT; 1941 AA.
AC Q9DGM4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE Fast myosin heavy chain isoform 3.

```

```

OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN 1;
RP SEQUENCE FROM N.A.
RA Zhang Q., Bandman E.;
RT "Seven skeletal myosin heavy chain genes (MyHC) are organized as a
RT multigene complex in the chicken genome.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF272034; AAF99315.1; --
DR HSSP; P13538; 2MYS.
DR GO; GO:0016459; C:myosin; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
SQ SEQUENCE 1941 AA; 223162 MW; 9C8597C1CCF1DEBC CRC64;

Query Match 9.6%; Score 260; DB 13; Length 1941;
Best Local Similarity 22.6%; Pred. No. 0.00019;
Matches 118; Conservative 82; Mismatches 181; Indels 142; Gaps 16;

Qy 33 SQAPKPEGAQRTAGSALRDVSELSQLEDILSTYCVNNOGPGEDGAQGEPAEPE 92
Db 1093 SIOQKIEDEQALGWQ---LQKKILEQARIEL-----EEIEAEITSRA 1135
Qy 93 DAEKRTYVARNGEPEPTPVVYGEKPSKGDNTBEIROSDEVGDRHRRPQKKAAGL 152
Db 1136 KAKKHRAADLSRELE-----ISERLEAGGATAAQI-----DNKKREAEFQVRRD 1182
Qy 153 GKEITLLMOTLNTLSTPEKLAALCKY-----AELEHNSQKQKMLLOKQSLVQEK 208
Db 1183 LEETL-----QHEATAAALRKCHADSTABIEGEQIDNLRVKQKLEKSELSAMEI 1233
Qy 209 DHLRGHSHKAVLARSKLESICREL-----QRNRSIKKEGVQAR----- 248
Db 1234 DLLASNMESVSKAKANLEKMCRTLEDQLESEVTKKEEHQRMINDLNTQARLTQEAEGYS 1293
Qy 249 -----EEERKKEVTSH-FQVTLNDIQLQWEOH 276
Db 1294 RVVEEKDALISLSRGKAFTQOIIEELKRLKEETKAKNALAHALQASARHDCDLREQYE 1353
Qy 277 ER-----NSKLR-----QENMELAERLKGLIEQVELREEHIDKVF 311
Db 1354 EEOEAKGELQALSKANSEVAQWRTKYETDAIQTREEEAKKLAQRLQDAEHEV----- 1409
Qy 312 KHKDLOQQQVDAKLAQAQAEMLKEAEERHOREKDFLKAEAVESQRMCELMKQOETHLKQOL 371
Db 1410 -----EAVNAKAS-----LEKTKQLQNEVEDLMDVERANAAACAALDKKQKDFKIL 1458
Qy 372 ALYTEKPEEPQNTLSKSEVFTFPKQEMKMTKKLKEKETMTWRSWESSNKALLEMA 431
Db 1459 AEWKQYEEQAELEASQKESRSLSSTELFKM-----KNAYEESLDHLQTLKRENKLQOEI 1514
Qy 432 EERTVDRKELEGLQVKIQRLEKLCRALQTERNDLNKRVDLSA 474
Db 1515 SDLT--EQIABGGKA-IHELEKVKQIQEKKSELQAALAEBA 1554

RESULT 31

```

O9DGM5	O9DGM5	PRELIMINARY;	PRT; 1944 AA.
ID	O9DGM5	SEQUENCE FROM N.A.	
AC	O9DGM5	01-MAR-2001 (TREMELrel. 16, Created)	
DT	O9DGM5	01-MAR-2001 (TREMELrel. 16, Last sequence update)	
DT	O9DGM5	01-OCT-2003 (TREMELrel. 25, Last annotation update)	
DE	O9DGM5	Fast myosin heavy chain isoform 2.	
OS	O9DGM5	Gallus gallus (Chicken).	
OC	O9DGM5	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	O9DGM5	Archosauria; Aves; Neognathae; Galliformes; Phasianinae;	
OC	O9DGM5	Gallus.	
OX	O9DGM5	NCBI_TaxID=9031;	
RN	O9DGM5	[1]	
RP	O9DGM5	SEQUENCE FROM N.A.	
RA	O9DGM5	Zhang Q., Bandman E.;	
RT	O9DGM5	"Seven skeletal myosin heavy chain genes (MyHC) are organized as a	
RT	O9DGM5	multigene complex in the chicken genome."	
RL	O9DGM5	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.	
DR	O9DGM5	EMBL; AF272033; AAF99314.1; -	
DR	O9DGM5	PIR; S24348; S24348.	
DR	O9DGM5	HSSP; P13538; 2MYS.	
DR	O9DGM5	GO; GO:0016459; C:myosin; IEA.	
DR	O9DGM5	GO; GO:0005524; F:ATP binding; IEA.	
DR	O9DGM5	GO; GO:0003774; F:motor activity; IEA.	
DR	O9DGM5	InterPro; IPR000048; IQ_region.	
DR	O9DGM5	InterPro; IPR001609; Myosin_head.	
DR	O9DGM5	InterPro; IPR004009; Myosin_N.	
DR	O9DGM5	InterPro; IPR002928; Myosin_tail.	
DR	O9DGM5	Pfam; PF00612; IQ; 1.	
DR	O9DGM5	Pfam; PF00063; myosin_head; 1.	
DR	O9DGM5	Pfam; PF02736; Myosin_N; 1.	
DR	O9DGM5	Pfam; PF01576; Myosin_tail; 1.	
DR	O9DGM5	PRINTS; PR00193; MYOSINHEAVY.	
DR	O9DGM5	ProDom; PD000355; myosin_head; 1.	
DR	O9DGM5	SMART; SM00015; IQ; 1.	
DR	O9DGM5	SMART; SM00242; MYSC; 1.	
DR	O9DGM5	PROSITE; PS50096; IQ; 1.	
SQ	O9DGM5	SEQUENCE 1944 AA; 223211 MW; 1547438220072D68 CRC64;	
Query Match 9.6%; Score 259; DB 13; Length 1944;			
Best Local Similarity 22.6%; Pred. No. 0.00022;			
Matches 118; Conservative 82; Mismatches 181; Indels 142; Gaps 16;			
Qy	33	SOAPRKEGAQARTAQSGALRDVSEELSQLIEDILSTYCVNNQGGPGGAGCEPAEPE 92	
Db	1096	SQIQSKIEDEQALGMQ---LQKKIKELQARIEEL-----EEIEARTSRA 1138	
Qy	93	DAEKSTYVARGEPPTVPVYGEKPSKGDPTNTEIRQSDVGDGRDHRPQKKAAGL 152	
Db	1139	KAEXRADLSRELEE-----ISRLEAGGATATQI-----DMNKKREAFQKMRD 1185	
Qy	153	GKEITLMTQTLNTLSTPEEKLAALCKKY---AELEHNSQKMKLQKQSLVQEK 208	
Db	1186	LEEAUTL-----QHEATAAARKKHADSTAEIGQIDNLQVKQKLEKSEKMEI 1236	
Qy	209	DHLRGHSHKAVIARSKLESLELREL-----QRNRSLEKBEVGQAR----- 248	
Db	1237	DDLASNMESVKANLEKMCRTLEDQSLKSEEHORMINDLSTQARLQTESGEYS 1296	
Qy	249	-----NSKLR-----BEEKKEVTSH-FQVTLNDIQLQMEQHN 276	
Db	1297	ROVEKDALISQLSRGKQAFQTOQIEHLKRLHEEIKAKNALAHQARHDCDILLREYE 1356	
Qy	277	ER-----NSKLR-----QENMELARLKKLIEQYELREBHIDKVF 311	
Db	1357	EEQKAGELQALSKANSEVAQMTKYETAQIQRTEELBEAKKLAQLQDAEHHV--- 1412	
Qy	312	KHKLOQOLVDKIQQAQEMKAEABRHQRKQFLLKAEVSRQCBELMKQOETHLQOL 371	
Db	1413	-----SANNKACAS-----LETKQRLQNEVEDLMDIVERANAAACARLDKQKQNFDKIL 1461	
Qy	372	ALYTEKEEFQNTLSKSEVPTTFKQEMKMTKKIKKLEKETTMYRWRWSSNKALLEWA 431	
Query Match 9.5%; Score 257; DB 13; Length 1943;			
Best Local Similarity 21.9%; Pred. No. 0.00027;			
Matches 118; Conservative 83; Mismatches 181; Indels 158; Gaps 17;			
Qy	33	SOAPRKEGAQARTAQSGALRDVSEELSQLIEDILSTYCVNNQGGPGGAGCEPAEPE 92	
Db	1096	SQIQSKIEDEQALGMQ---LQKKIKELQARIEEL-----EEIEARTSRA 1138	
Qy	93	DAEKSTYVARGEPPTVPVYGEKPSKGDPTNTEIRQSDVGDGRDHRPQKKAAGL 152	
Db	1139	KAEXRADLSRELEE-----ISRLEAGGATATQI-----DMNKKREAFQKMRD 1185	
Qy	153	GKEITLMTQTLNTLSTPEEKLAALCKKY---AELEHNSQKMKLQKQSLVQEK 208	
Db	1186	LEEAUTL-----QHEATAAARKKHADSTAEIGQIDNLQVKQKLEKSEKMEI 1236	
Qy	209	DHLRGHSHKAVIARSKLESLELREL-----QRNRSLEKBEVGQAR----- 248	
Db	1237	DDLASNMESVKANLEKMCRTLEDQSLKSEEHORMINDLSTQARLQTESGEYA 1296	
Qy	249	-----NSKLR-----BEEKKEVTSH-FQVTLNDIQLQMEQHN 276	
Db	1297	ROADKKGGLISQLSRGKQAFQTOQIEHLKQLEEEIKAKNALAHLSQARHDCDILLREYE 1356	
Qy	277	ER-----NSKLR-----QENMELARLKKLIEQYELREBHIDKVF 311	
Db	1357	EELEKAGELQALSKANSEVAQMTKYETAQIQRTEELBEAKKLAQLQDAEHHVE--- 1413	

129 IRQDEVCDRRHRRPOEKAKGLGKBITLLMQLTNTLSTPEEKLAALCKKYAELLBEHR 188
 76 ERPAZI--KPPKPEERTAPSNLEAL-----ERKAKENKLEDEWR 119
 189 NSQKMLKQKQSQVORVOKHDLRGHSHKAVLAKSLKESLRELQHR--NRSLKEBGVQ 245
 120 N-----KAKSAMGERDLRSEIKR---LKEELEKOEKELDKYIKISKOLKEX-LE 165
 246 RAREEEKRKBVTSHFOVTLNDIQLQMEOHNRNSKLR---QENYELAERLKLIRQY-B 301
 166 KAKRESELEKAKAEYERVEKI---AGKNELKSLKLESLDQNRRLAENLKLKKEKINE 222
 302 LREHIDRVFKHDLQOLVDKILQQAQEMLEKAEERHOREKDFLLKAEVESORMCELMK 361
 223 IKEER-DLKEETKEVGLKD-QLAQLQSLKVEKS---ERDDLANEVEALRNEKLR 276
 362 QOETHLKOQALYTEKEPEEFQNTLSKSSEVFTTTPKQEMKVTXKIKLEKETTMYR---S 418
 277 KIDKLSAESUNLOKLDREKLEKAKOHIGKLRBEIKRDRBEIKRKAQSKLDEIK 336
 419 RWESNKKALL-----EMAEKTVDRKLEGLQVKIQRLEK 453
 337 RYEBGKLLVPPEEMALIAK-VKGSIVIGKNSMVKALQE 374

RESULT 35

O08815 ID O08815 PRELIMINARY; PRT; 1206 AA.
 AC O08815;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Protein kinase.
 GN SK2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SD;
 RA Fukami Y., Yamamoto H., Ichihara T., Mori K., Gomi T., Sato K.;
 RT "SK2, a putative rat homologue of yeast protein kinase NRK1";
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SD;
 RX MEDLINE=91287726; PubMed=2062320;
 RA Yue C.C.;
 RT "Novel putative protein kinase clones from a rat large granular
 lymphocyte tumor cell line.";
 RL Mol. Immunol. 28:399-408(1991).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AB003357; BAA20077.1; -;
 DR FIR; FT0204; FT0204.
 DR PIR; T34021; T34021.
 DR HSSP; P24941; 1CKP.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0004518; F:nuclease activity; IEA.
 DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006289; F:nucleotide-excision repair; IEA.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
 DR InterPro; IPR000719; Prot kinase.
 DR InterPro; IPR002290; Ser thr pkinase.
 DR InterPro; IPR008271; Ser thr pkinase.
 DR InterPro; IPR001943; UvrH/C.
 DR Pfam; PF00069; pkinase; 1.
 DR SMART; PD000001; Prot kinase; 1.
 DR SMART; SM00220; S TKC; 1.
 DR PROSITE; PS00107; "PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS0151; UVR; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 1206 AA; 137888 MW; 8P5C6A19A3501PB9 CRC64;
 Query Match 9.4%; Score 253; DB 11; Length 1206;
 Best Local Similarity 22.2%; Pred. No. 0.00026;
 Matches 133; Conservative 92; Mismatches 199; Indels 176; Gaps 25;
 QY 12 EGQRPSPQAAVAERAGPGSSQA-----PR 37
 DB 682 ENAQELFVKAEP-----QAPASQASEPPVLPISINIHSENTENKEMALPKPETILPP 737
 QY 38 KPEGAQARTAQSG--ALADVSEELSRQLEDILSTYCVDNNGQGGPDCDAQEPARPEDA 94
 DB 738 EPENGKNDTSGTGSTVENSDDLNLISFSL-----KTKDGSVSLQETRRQKTL 791
 QY 95 EKSRTVVARNGPEPTPVYGEKFSKGPNTETIRQSDVGDRRRRRQEKKAKGLCK 154
 DB 792 KXTRKPIVDGVSVVT---SKIVTDSKTEELR-----FLRQELRELELQK 838
 QY 155 EITLLMQTLN-TLSTPEEKLAALCKKYAELLBEHNSOKMKLLQKQQLVQEKDHLRG 213
 DB 839 EEQKAOQQLNGKLQOOREQIFRRFEQ--EMLSKKQYDOEIELEKQKQTIE-----889
 QY 214 EHSKAVLARSKLESICRELQRNRSIKKEGVQARAEERKEKVTSHFOVTLNDIQLQME 273
 DB 890 -----RLE-----QETNRLDE-AKRIKGEQKE---LSKFQNLNRN-----923
 QY 274 QHNERNKLRQENMELAEKLLISQ-----YELREEH 306
 DB 924 RKKEGEFVKQKQQLDGLKIIQQKAKLANIERECLNTQQLLRAREAAIWELEERH 983
 QY 307 IDKVPKHDLOQOLVDKILQQAQEMLEKAB---ERHOREKDFLLKAEV--ESQRMCELMK 361
 DB 984 LQE--KHOLLKQQLKDQVFIQRHLLKREKETEQQRYNQRLIBELKNRQTERARLPK 1041
 QY 362 QOETHLKOQALY-----TEKEPEEFQNTLSKSSEVFTTTPKQEMKVTXKIK 406
 DB 1042 IQRSKAKTNAMPKSKSLRINSTATPDQREKIKQFAAQSEK-----RQNERMAQH- 1092
 QY 407 KKLKETTMYRSRWBSSNKALAEAEK-----TVDRKELE---GLQVKIQRLEKLC 455
 DB 1093 QKHSQMRDLQOCERANVRLEHLQNEKCHLVAHEHTQKLELDEHSEQLKEWR-EXLR 1151
 QY 456 RALQTERNDLNRVQDLS-----AGQGSLLTDSGPERPEGPGQAQAPSPRTEAPCYPG 510
 DB 1152 PRKKTLEEFARFKQEQVFPFMTGESECLNPSAQSR-----GCLQTHSPSSTRAPAWAG 1206

RESULT 36

O8AY28 ID O8AY28 PRELIMINARY; PRT; 1940 AA.
 AC O8AY28;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Fast myosin heavy chain HCIII.
 GN N127.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhang Q.J., Bandman E.;
 RT "Seven Skeletal Myosin Heavy Chain Genes (MyHC) are Organized as a
 Multigene Complex in the Chicken Genome.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY116218; AAM88910.1; -;
 DR GO; GO:0016459; C:myosin; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.

[illegible]

QY	399	MERMTHKIKKLEKETTMYRSRWSSNKKALLEMAEKTVDKLEGLQVKIQRLKCRAL	458
Db	1155	VEHETQKLEDEHSHQELKWR-----EKLPRKTYL-----EFPARKL	1195
QY	459	QTE 461	
Db	1196	QEQ 1198	
RESULT 38			
ID	Q07569	PRELIMINARY; PRT; 2139 AA.	
AC	Q07569; O02504;		
DT	01-NOV-1996 (T-EMBLrel. 01, Created)		
DT	01-JUL-1997 (T-EMBLrel. 04, Last sequence update)		
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)		
DE	Myosin heavy chain.		
GN	MECA.		
OS	Entamoeba histolytica.		
OC	Eukaryota; Entamoebidae; Entamoeba.		
OX	NCBI_TaxID=5759;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=HMI:IMSS;		
RX	MEDLINE=93295430; PubMed=8515774;		
RA	Raymond-Denise A., Sansonetti P., Guillen N.;		
RT	"Identification and characterization of a myosin heavy chain gene		
RT	(mhcA) from the human parasitic pathogen Entamoeba histolytica.";		
RL	Mol. Biochem. Parasitol. 59:123-131 (1993).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=HMI:IMSS;		
RA	Guillen N.;		
RL	Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; L03534; AAB48065.1; -.		
DR	PIR; T18296; T18296.		
DR	HSSP; P08799; 1WMD.		
DR	GO; GO:0016459; C:myosin; IEA.		
DR	GO; GO:0005524; F:ATP binding; IEA.		
DR	GO; GO:0003774; F:motor activity; IEA.		
DR	InterPro; IPR000048; IQ region.		
DR	InterPro; IPR001609; myosin_head.		
DR	Pfam; PF00612; IQ; 2.		
DR	Pfam; PF00063; myosin_head; 1.		
DR	PRINTS; PR00193; MYOSINHEAVY.		
DR	ProDom; PD000355; myosin_head; 1.		
DR	SMART; SM00242; MYSC; 1.		
SQ	SEQUENCE 2139 AA; 245225 MW; C68307341DB51DD1 CRC64;		
Query Match			
Best Local Similarity 9.2%; Score 249.5; DB 5; Length 2139;			
Matches 106; Conservative 93; Mismatches 172; Indels 117; Gaps 19;			
QY	54	DVSELSRQLDEILSTYC-VNNMGSGGEGGAQGE-----PAPPE 92	
Db	1169	DVSQYLRQKEEYSQIAKMQEKEAIGNDVYKNEKTIKELBIQSLQKLDTEYEKE 1228	
QY	93	DAEKSRTVVRNGEPPTPVVYVGEKSPKDPNTEIRQSDVGDROHRRPOEKKAKGL 152	
Db	1229	DAEKKKEIEKE-----MKALQKEKE-----NVESKKNSTE-----KDKKKLEDN 1268	
QY	153	GKEITLMTNTISTPEKLAALCK-----KYALLEHRSNQMKLLOKQSLVQEK 208	
Db	1269	LKD---TQKLDMDTADNEKLKAKAKDLEAQINEVQNHKAVADAEILANKKKAQSDKEL 1325	
QY	209	DHLRGEHSAVLARSLKESLCRELQRNRSLKESGVOPAREEEKKEKTEVTSFOVTINDI 268	
Db	1326	NSLKAEALTKASVSKNSKQNEKAALSEE-IDQNEK-----LKNIQADLRKA 1377	
QY	269	QLQMEHNRNLSKQENNELA---ERLKLIEQVELE-----BHIDVFKHDLQ 318	
Db	1378	TADLQAEANEKAEVBAQRDLVADNKKWTKLEIKARDEENTYKVENYKVLKCK--- 1433	

QY	319	QLVDAKLAQAQAEMLK-BAEERHQREKDFLLKEAVESORMCELMKQOETHLQQLA----	L 373
Db	1434	---EADLEAEANENLDIEKKDRMKEQV---KLEGE-----LKETDKLAAIAEKDSI	1482
QY	374	YTEK-----PEEPQNTLSKSSEVPTTFKQENKMTKKIKLEKETTYVRSWESSNKAL	427
Db	1483	PTAKKQSDADLESLNKTVEEHDEVVAKLTGITKLTRDNQSAABELNELRSKADKKKKI	1542
QY	428	LEMAE-----EKTVRD-----KELEGVQVKIQRLKCRALQTER 462	
Db	1543	SELEQVNELSRPVGTGTVADENEIKINDAQIADLANKALEMKGVQNNQQAATNKELKAKD	1602
QY	463	NDLNRKVQ 470	
Db	1603	NDLTSKIE 1610	
RESULT 39			
ID	Q9CS72	PRELIMINARY; PRT; 610 AA.	
AC	Q9CS72;		
DT	01-JUN-2001 (T-EMBLrel. 17, Created)		
DT	01-JUN-2001 (T-EMBLrel. 17, Last sequence update)		
DT	01-OCT-2002 (T-EMBLrel. 22, Last annotation update)		
DE	5730485H2IRik protein (Fragment).		
GN	5730485H2IRIK		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Embryo;		
RX	MEDLINE=21085860; PubMed=11217851;		
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,		
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,		
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,		
RA	Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,		
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,		
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,		
RA	Nordone P., Ring B., Ringwald W., Rodriguez I., Sakamoto N.,		
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,		
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,		
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohetsuki S.,		
RA	Hayashizaki Y.;		
RT	"Functional annotation of a full-length mouse cDNA collection.";		
RL	Nature 409:585-590 (2001).		
DR	EMBL; AK017709; BAB30888.1; -.		
DR	MGI; MGI:1917848; 5730485H2IRik.		
DR	InterPro; IPR000533; Tropomyosin.		
DR	PRINTS; PR00194; TROPOMYOSIN.		
FT	NON TER 610		
SQ	SEQUENCE 610 AA; 70346 MW; 9B9EFC87F64PCB80 CRC64;		
Query Match			
Best Local Similarity 9.2%; Score 249; DB 11; Length 610;			
Matches 129; Conservative 86; Mismatches 217; Indels 98; Gaps 21;			
QY	10	GP-EGAEQPSQAAPAVEAEPGSSQAQPRPEGAQARTAQSGALRDVSELSRQLEILS 68	
Db	31	GPSADAKKNAKNGKGGDDVMASGTVKRHLKPSGESEKTKKPL-----ELSK--EDLIQ 82	
QY	69	TYCV-----DNNQGGPDGAGQEPAPEDAEKRTTYVARGEPPTPVVYGEKE 118	
Db	83	LLSIMEGELQVGRNRSNQ--GSSSNHGVSCP-----SISSDGGKGPSEDAKKNKA 135	

```
QY 119 PSKGDNTBIROSDEVGRDRHRP-----QEKKAKGLCKS--ITLLMQLTNTLSTPRE 171
Db 136 NRKGE---DDVWASGV--XHLKPGSEKTKKPLESKEDLIQLLSIMSELQARHD 190
QY 172 KLAALC--KKYABELBEHRNSQOMKLLQ--KKOSOLVOKHDLRGEHSAVLARSKLS 227
Db 191 VHMMLTEKTKPVLGAHYSAPPEKVLRLVLRDAILAQEKSGEDVVEKPISELDRLEE 250
QY 228 LCHELOR---HNSLKEEGVQARAREEKEKEVTSHFQVTLNDIQLQMHQHRNSKLRQ 284
Db 251 KQKETVRMLBQLLLAEKCHRRTVYLENEKHKHTDYNNKSDFTNLLQERLERLKLLE 310
QY 285 B-----NMELAEKLLIEQ-----YELREHIDKVKHDLQQLVD--- 322
Db 311 QEKAYOARKEKENAKRLNKLRLDELVKLSPALMLVDERQWHIEQL---GLSQKQVODLT 366
QY 323 AKLQQAQEMLKAEERHOREKDFLLKEAV-----ESQRYCELMKKQETHLQO--- 369
Db 367 QKURSEEEKKALITYSKEDQKLLKLEVDPEHKASRFQOEHEEMNAKLANGESHNRQUR 426
QY 370 ----QLALYTEKEPFTNTLSKSEVFTTFKQEMERTKKIKKLEKETTYMSRWESSNK 425
Db 427 LKLVGLSQRIEELEETNKSQKAEB-----ELQELRDKIAKGECSNLSMAEVESLRK 479
QY 426 ALLEMA---BEKTVRKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDL 472
Db 480 RVLEMEGQDEBITKTAQCRKELKQBEHHSKELRLEVEKLCQKRMSEL 529

RESULT 40
Q91LX9 PRELIMINARY; PRT; 1003 AA.
AC Q91LX9
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE ORF73.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=37296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21376412; PubMed=11483733;
RA Garber A.C., Shu M.A., Hu J., Renne R.;
RT "dna binding and modulation of gene expression by the latency-
RT associated nuclear antigen of Kaposi's sarcoma-associated
RT herpesvirus.";
RL J. Virol. 75:7882-7892(2001).
DR EMBL: AF360120; AAK50002.1; -.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR000533; Tropomyosin.
DR PRINTS: PR00194; TROPOMYSIN.
SQ SEQUENCE 1003 AA; 115517 MW; C20C43308B01A0A3 CRC64;

Query Match 9.2%; Score 249; DB 12; Length 1003;
Best local similarity 20.0%; Pred. No. 0.00033;
Matches 103; Conservative 108; Mismatches 217; Indels 86; Gaps 13;

QY 11 PEGAQERPSQAAPVAEAGGSGQAQRPK-----EGAAQ-RTAQSGALRDVS 56
Db 236 PFWATESPIYVGSSDGDTP-----PRQPTSPISIGSSPSGSGMGDDTAMLVLLAEIA 290
QY 57 HELSRQLEDILTSLCYDNNQGG-----PGEDCAQGEPAPEPAEAKS 97
Db 291 HEASKNEKE-----CSENQAGEDGDNELSKESQVYKDDNKNKDEEEQETDEDEDD 345
QY 98 RTYVANGE-----PETPVVYGEKPSKGDPNTEEIRQSGDEVGRDRHRP 143
Db 346 EKKDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDE 405
QY 144 QEKKAKK---GLGKKTITLLMQLTNTLSTPREKLAALCKKYALLBEHRNSQOMKLLQK 199
```

```
Db 406 BEDKKEDESDGDGNK-TLSIQSSQQQEPQO-----QBPOQEPQOQBPOQEPLOE 457
QY 200 KQSOLVOKHDLRGEHSAVLARSKLSLCELRHARSLKKEGCVQARAREEKEKEVT 259
Db 458 PQQEPPOQEPPOQEPQO-----EPQOQEPPOQEPPOQEPPOQEPPOQEPPOQEP 510
QY 260 HFQVTLNDIQLQME-QHNERNSKLRQNMELAEKLLIEQYELREHIDKVKHDLQO 318
Db 511 QDEQEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQ 569
QY 319 QLVDAKLAQQAQEMLKAEERHOREKDFILKEAVESQRMCHLMKKQETHLQQLALYTEKF 378
Db 570 QELREHIDKVKHDLQOQLVDEPEHKASRFQOEHEEMNAKLANGESHNRQUR 621
QY 379 REFQNTLSKSEVFTTFKQEMERTKKIKKLEKETTYMSRWESSNKALLENABEKTVRD 438
Db 622 EBOEELREHIDKVKHDLQOQLVDEPEHKASRFQOEHEEMNAKLANGESHNRQUR 681
QY 439 KELEGLOVKIQRLEKLCRALQTERNDLNKRVQDL 472
Db 682 QELREHIDKVKHDLQOQLVDEPEHKASRFQOEHEEMNAKLANGESHNRQUR 715

RESULT 41
Q9P216 PRELIMINARY; PRT; 1005 AA.
AC Q9P216
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hypothetical protein KIAA1361 (Serine/threonine kinase TA01)
DE (fragment).
GN KIAA1361.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:165-73(2000).
RN [2]
RP SEQUENCE OF 5-1005 FROM N.A.
RA Jenkins S.G., D'Andrea R.J., Gamble J.R., Vadas N.A.;
RT "Characterization of human TA01.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: AB037782; BA032599.1; -.
DR EMBL: AY049015; AAL12217.1; -.
DR HSP; P24941; I338.
DR GO: GO:0005524; P:ATP binding; IEA.
DR GO: GO:0004674; P:protein serine/threonine kinase activity; IEA.
DR GO: GO:0004674; P:transferase activity; IEA.
DR GO: GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro: IPR000719; Prot kinase.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR008271; Ser_thr_kinase.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Prot kinase; 1.
DR SMART: SM00220; S_TKc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW Hypothetical protein; ATP-binding; Kinase;
KW Serine/threonine-protein kinase; Transferase.
FT NON_TER 1
SQ SEQUENCE 1005 AA; 116540 MW; A83BD4E53569BCB5 CRC64;
```

Query Match 9.2%; Score 248.5; DB 4; Length 1005;
 Best Local Similarity 21.6%; Pred. No. 0.00035;
 Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

QY 23 PAVEAE-----GPGSSQA-PRKPEGAQARTAQSGALRDVSELSRQLE 64
 DB 328 PAVEAQEEEEEDHGVRGTGTWNSVGSNQSIIPMSISASSQSSVNSLPDVSDDKS-ELD 386
 QY 65 DILSTYCVNNQ-----GPGEDG--AQGEPAEPEDAESRTTVARNGEPEPTPVYGEKE 118
 DB 387 NMEGDHTVMSNSVIHLKPEBENTREGDP-----RT---RASDPQSPPOVSRHKS 434
 QY 119 PSKGDPTNEERQSDDEVDRHRPQEKKAAGLKEIT-----LLMQTLNTLSTPEEKLA 174
 DB 435 HYRNRHPTATITASLV---TRQMEHQDSSELREQSGYKMRQRQKQMLTLENKLA 490
 QY 175 ALCKYA---ELIEHRN---SQKMKLLQKQSQVQV-----KDHLRGHSK 217
 DB 491 AEMDEHRLRLDKLETQRNFPAAEKELIKHQHAMEKBAKVMNEEKKKFOQHQAQCKK 550
 QY 218 AVLARSKLESICRELQRHNRSLKEGVOR---AREBEE---KRKEVTSHPV---TLNDI 268
 DB 551 EL--NSPLESQKREYKLRKEQLKEELNENQSTPKKEQEWLSKQENIQHFQABEEANLL 608
 QY 269 QLOWE-----QHN-----ERNKLRQENMELARLKKLIEQYELRE 304
 DB 609 RRQRQYLELCRRFRKRRMLLGRHNLQDLVREELNKRQTQKDLHAWMLLRCHESMQBLEF 668
 QY 305 EHIDKVFHKDLQOOLVDAKLOQAQEMLEKAEERHOREKDFLLKEAVESQRMCELMKQOE 364
 DB 669 RLNTIQK---MRCELI---RLQHQTELNTQLEYKRRERELRRKHVMEVQOPKSLASKE 723
 QY 365 THLQOOL---ALYTEKPEEPONTLSKSSSVFTTFQKQMEKTKKIKKLEKETTMYRSRW 420
 DB 724 LQIKKQFDQTKIQTRQYKALRNHLE---TTPKSEHKAVLKRLK--BEQTRKLAIAE 776
 QY 421 ESSKALLEMAEKTVR-----DKELE 442
 DB 777 EYDHSINEMLSQALRLDRAQAEACQVLKMLQLOLELLNAYOSKIKMQABEAQHRELRL 836
 QY 443 GLOVKI-----QRLEKLCRALQTERND-----LNKRVDLSAGQGS 480
 DB 837 ELEQVSLRRALLQKIEEEMALQNERTERIRSLERQAREIEAFDSMRIGFSNNWL 896
 QY 481 TDSQPERPEPGAQAPSSPVTEAPCYGPAP---STEASQGTGP 522
 DB 897 SNLSPE-----AFSHSYPGASSWSHNPCTGCGPGP 924

RESULT 42
 DB 088664 PRELIMINARY; PRT; 1001 AA.

AC 088664;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Serine/threonine protein kinase TA01.
 CS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9903202; PubMed=9786855;
 RA Huchison M., Berman K.S., Cobb M.H.;
 RT "Isolation of TA01, a protein kinase that activates MEKs in stress-
 activated protein kinase cascades.";
 RL J. Biol. Chem. 273:28625-28632(1998).
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AF084205; AAC71014.1; -
 DR PIR; T17365; T17365.
 DR HSP; P24941; 1B38.
 DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
 DR InterPro; IPR007019; Prot kinase.
 DR InterPro; IPR002290; Ser thr pkinase.
 DR InterPro; IPR008271; Ser thr pkin_AS.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot kinase; 1.
 DR SMART; SM00220; S_TKc_1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 1001 AA; 115952 MW; 85511B62DBD62FCC CRC64;

Query Match 9.1%; Score 247; DB 11; Length 1001;
 Best Local Similarity 21.3%; Pred. No. 0.00041;
 Matches 142; Conservative 118; Mismatches 212; Indels 176; Gaps 30;

QY 23 PAVEAE-----GPGSSQA-PRKPEGAQARTAQSGALRDVSELSRQLE 64
 DB 324 PAVEAQEEEEEDHGVRGTGTWNSVGSNQSIIPMSISASSQSSVNSLPDASDDKS-ELD 382
 QY 65 DILSTYCVNNQ-----GPGEDGACGEPAEPEDAESRTTVARNGEPEPTPVYGEKPS 120
 DB 383 NMEGDHTVMSNSVIHLKPEBENYO-BEGDPR-----TRASAPQSPPOVSRHSHY 432
 QY 121 KGDPTNEERQSDDEVDRHRPQEKKAAGLKEIT-----LLMQTLNTLSTPEEKLAAL 176
 DB 433 RNREHPTATITASLV---TRQMEHQDSSELREQSGYKMRQRQKQMLTLENKLAKE 488
 QY 177 CKYA---ELIEHRN---SQKMKLLQKQSQVQV-----KDHLRGHSKAV 219
 DB 489 MDEHRLRLDKLETQRNFPAAEKELIKHQHAMEKBAKVMNEEKKKFOQHQAQCKEL 548
 QY 220 LARSKLESICRELQRHNRSLKEGVOR---AREBEE---KRKEVTSHPV---TLNDIQL 270
 DB 549 --NSPLESQKREYKLRKEQLKEELNENQSTPKKEQEWLSKQENIQHFQABEEANLLR 606
 QY 271 QME-----QHN-----ERNKLRQENMELARLKKLIEQYELREEH 306
 DB 607 QRQYLELCRRFRKRRMLLGRHNLQDLVREELNKRQTQKDLHAWMLLRCHESMQBLEFRH 666
 QY 307 IDKVFHKDLQOOLVDAKLOQAQEMLEKAEERHOREKDFLLKEAVESQRMCELMKQOETH 366
 DB 667 INTIQK---MRCELI---RLQHQTELNTQLEYKRRERELRRKHVMEVQOPKSLKELQ 721
 QY 367 LKQOL---ALYTEKPEEPONTLSKSSSVFTTFQKQMEKTKKIKKLEKETTMYRSRWES 422
 DB 722 IKKQFDQTKIQTRQYKALRNHLE---TTPKSEHKAVLKRLK--BEQTRKLAIAEQ 774
 QY 423 SNKALLEMAEKTVR-----DKELEGL 444
 DB 775 YDHSINEMLSQALRLDRAQAEACQVLKMLQLOLELLNAYOSKIKMQABEAQHRELREL 834
 QY 445 QVKI-----QRLEKLCRALQTERND-----LNKRVDLSAGQGS-----LTD 482
 DB 835 EQVSLRRALLQKIEEEMALQNERTERIRSLERQAREIEAFDSMRIGFSNNVLN 894
 QY 483 SGPERPEPGAQAPS-SPVTEAP-CYGPAPSTEASG---QTGQP 523
 DB 895 LSPFAFHSYPGASSWSHNPCTGSGPHWGHMGTPQAWGHPMQCGPQ 942

RESULT 43
 DB 0875C7 PRELIMINARY; PRT; 2055 AA.

AC 0875C7;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Chimeric erythrocyte-binding protein MABBL.
 GN MABBL.

```

OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Blais B.L., Kappe S.H.I., Maciel J.S., Balu D.J., Adams B.;
RT "Plasmodium falciparum maeli is a unique member of the ebl family.";
RL Mol. Biochem. Parasitol. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Michon P., Stevens J.R., Kaneko O., Adams J.H.;
RT "Plasmodium falciparum maeli is a unique member of the EBL family.";
RL Mol. Biol. Evol. 0:0-0(2002).
RN [3]
RP SEQUENCE FROM N.A.
RA Chai M., Dutta S., Ockenhouse C.F.;
RT "Identification, expression and functional characterization of MAEBL,
RT a sporozoite and asexual blood stage chimeric erythrocyte binding
RT protein of Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 0:0-0(2002).
DR EMBL; AF042084; AAL10509.1; -.
DR EMBL; AF400002; AAM90625.1; -.
DR HSSP; P04286; IIC2.
DR InterPro; IPR008602; Duffy binding.
DR Pfam; PF05424; Duffy binding; 1.
SQ SEQUENCE 2055 AA; 243210 MW; 783P48BB7286B1F3 CRC64;

Query Match          9.1%; Score 246.5; DB 5; Length 2055;
Best Local Similarity 24.5%; Pred. No. 0.0009;
Matches 101; Conservative 85; Mismatches 146; Indels 81; Gaps 16;

QY 116 EKPSKGDPTNTTEIRQSDV-----GDRHRRPQEKKAKGLGKEITLLMTLNTL 166
DB 1436 EKKAENLKAAEKKKADLKKAEKKKADLKKAEKKKADLKKAE--EKKADEL 1494
QY 167 STPEKLAALCKKYAEL--LEHRSNQKMLQKQSQVQKDLHGEHSHKAVLARSK 224
DB 1495 KKAEEK-----KKADELKKAEEKKADLKKAEKKKADLKKAEELKKAEEK-----KK 1544
QY 225 LESLRELQHRNRSIKE-----EGVQRAEEREEKKEVTSHFQVTLNDIQLQM 272
DB 1545 VQKKREERERNMARRAAILKQIEKKRIEVMKLYEKKKAE-----QLKK 1593
QY 273 EQHNRNSKLRQENMELARLKLHGYELREHHDVFKHKLQOOLVDKLAQQAQEML 332
DB 1594 EE-----EEKIKAEQKKEEEKKQVQKKEEEKK--KAEQKKEEENKIK--AEOL 1645
QY 333 KKAERHQRKDFLLKAEVSRMCELMKQOETHLK--QOLALYTEKPEEFQNTLSKSE 390
DB 1646 KKAEEKKKAEELKKEEEKKKAEQKKEEEKKQVQKKEEEKKAEQKKEE 1705
QY 391 ----VFTTFQEMERTYKIKKLEKETTMYRSRWSSNKALLEMAPEKTVRDKLEGLQV 446
DB 1706 ENKIKVEQLKKEEEKKAEELKKE-----EEKKKVQQLKKE---EKKAEEIRK 1754
QY 447 KIQRL--EKLCRALQTERNDLNRVQDLSAGGSGSLTDSGPERRPGGAQAP 497
DB 1755 EKAAVEIEELKKDEKGRMEVEKKIKD-----TKDNFENIQEGNNKNTP 1798

RESULT 44
Q8IHP3
ID Q8IHP3 PRELIMINARY; PRT; 2055 AA.
AC Q8IHP3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MAEBL, putative.
GN Pf11.0486.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12369864;
RA Carlson J.M., Hall N., White O., Berriman M., Hyman R.W.,
RA Gardner M.J., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow J.J., Suh B., Peterson J., Angiuoli S.,
RA Petrea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.S.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 415:498-511(2002).
DR EMBL; AF014843; AAN36086.1; -.
DR InterPro; IPR008602; Duffy binding.
DR Pfam; PF05424; Duffy binding; 1.
SQ SEQUENCE 2055 AA; 243210 MW; 783P48BB7286B1F3 CRC64;

Query Match          9.1%; Score 246.5; DB 5; Length 2055;
Best Local Similarity 24.5%; Pred. No. 0.0009;
Matches 101; Conservative 85; Mismatches 146; Indels 81; Gaps 16;

QY 116 EKPSKGDPTNTTEIRQSDV-----GDRHRRPQEKKAKGLGKEITLLMTLNTL 166
DB 1436 EKKAENLKAAEKKKADLKKAEKKKADLKKAEKKKADLKKAE--EKKADEL 1494
QY 167 STPEKLAALCKKYAEL--LEHRSNQKMLQKQSQVQKDLHGEHSHKAVLARSK 224
DB 1495 KKAEEK-----KKADELKKAEEKKADLKKAEKKKADLKKAEELKKAEEK-----KK 1544
QY 225 LESLRELQHRNRSIKE-----EGVQRAEEREEKKEVTSHFQVTLNDIQLQM 272
DB 1545 VQKKREERERNMARRAAILKQIEKKRIEVMKLYEKKKAE-----QLKK 1593
QY 273 EQHNRNSKLRQENMELARLKLHGYELREHHDVFKHKLQOOLVDKLAQQAQEML 332
DB 1594 EE-----EEKIKAEQKKEEEKKQVQKKEEEKK--KAEQKKEEENKIK--AEOL 1645
QY 333 KKAERHQRKDFLLKAEVSRMCELMKQOETHLK--QOLALYTEKPEEFQNTLSKSE 390
DB 1646 KKAEEKKKAEELKKEEEKKKAEQKKEEEKKQVQKKEEEKKAEQKKEE 1705
QY 391 ----VFTTFQEMERTYKIKKLEKETTMYRSRWSSNKALLEMAPEKTVRDKLEGLQV 446
DB 1706 ENKIKVEQLKKEEEKKAEELKKE-----EEKKKVQQLKKE---EKKAEEIRK 1754
QY 447 KIQRL--EKLCRALQTERNDLNRVQDLSAGGSGSLTDSGPERRPGGAQAP 497
DB 1755 EKAAVEIEELKKDEKGRMEVEKKIKD-----TKDNFENIQEGNNKNTP 1798

RESULT 45
Q9WU41
ID Q9WU41 PRELIMINARY; PRT; 1202 AA.
AC Q9WU41;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Ste20-related kinase SMAK.
GN STK2 OR SMAK.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Sabourin L.A., Seale P., Tamai K., Wagner J., Rudnicki M.A.;
RT "Caspase 3 cleavage of the Ste20-related kinase SLK releases and
RT activates an apoptosis-inducing kinase domain and an actin-
RT disassembling region."

```


RESULT 47
Q80U65 PRELIMINARY; PRT; 1307 AA.
AC Q80U65;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE MKIAA0204 protein (Fragment).
GN MKIAA0204.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA genes:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
DR EMBL; AK122218; BAC65500.1; -;
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004518; F:nuclease activity; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0006289; F:nucleotide-excision repair; IEA.
DR GO; GO:0006468; F:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser Thr kinase.
DR InterPro; IPR008271; Ser Thr pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR001943; UvrB/C.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00151; UVR; 1.
FT NON TER 1
SQ SEQUENCE 1307 AA; 148738 MW; 61178A1329DDCA24 CRC64;

Query Match 9.0%; Score 244; DB 11; Length 1307;
Best Local Similarity 21.2%; Pred. No. 0.00075;
Matches 133; Conservative 99; Mismatches 183; Indels 212; Gaps 28;

QY 6 QPEAGPE--GAQRP-----SOAAPAV-EAEG--PGSQAPRKPEGAQARTASG 50
DB 687 KTEGPEAGAEHPGGRVEDKQPEQPAVCAEAGQLTSETTR---ATLEQPETD 742

QY 51 ALRDVSELS-RQLEDILTYVDNNOGPGEDG-----AQ 85
DB 743 EYEQVSESNISIELELVVTGAERALGSEGAATEVDLERKENAQVPVKAESQAPAA 802

QY 86 GPAPAEPE-----DAKSTTYVARNGEPPTPVYGEKPSKG---DPNT----- 126
DB 803 SQPSEHPVLIPSIINSETTENKEEMGALPKPEILPPPEHEKGNUTDSTGTSVENS 862

QY 127 -----EIRQSEVGDGRDHRPQ--KKAKGL--GKEITLIMQTLNTLSTPE 170
DB 863 SGDLNLSSFLSKANDSGSVLSQETTRQKTLKTRKPFVDGVEVSV---TTSKIVTDS 919

QY 171 EKLAALCKVYALLEHRNSQOMKLLQKQ-----SOLVQEKDHLRGEHSKAVLAR 222
DB 920 D-----SATEELRFLRRELRLQKEORAOQQLNGKLOQOQREQIFRRFEQEMLSK 973

QY 223 SK-----LESCLRELQRHNRSLKRGVORAREBE-----EKKEVTSHFOVTLNDIQLQM 272
DB 974 KFOYDOEIEENLEKQKQKOTIERLEQSHTNRLRDEAKRIKGEQEKEL-SKFNVLKN----- 1027

QY 273 EQHNRNSKLRQENNELAERLKLIEQ-----YELREE 305
DB 1028 -RKKEEQEFVQKQQLDGLSLKIIQQCAELANIERECLNNKQQLMRAREAAIWELEER 1086

QY 306 HIDKVFKKDLQOQLVDAKLOQAQENLKEAE---ERHOREKDFLKEAV--BSQRCMLM 360
DB 1087 HLQE--KQLLKQQLKQVFMQRHQLLKHKEKETQMQRYNQRLIEELKNRQTOERARLP 1144

QY 361 KQETHLQKQALY-----TEKPEFP---ONTLSKSSEVFTTFKOE--- 398
DB 1145 KIQSEAKTRMAMFKSLRINSTATPDQDREKIQFAAQBEKQKQVMAHQKHESQNR 1204

QY 399 -----MEQMTKKIKLEKETTMYRVSWSNKKALLEMAEBK 434
DB 1205 DLQQLCEANVRELHQLQNEKCHLLVEHETQKLDELDEHSOELKEWR-----EKL 1254

QY 435 TVRDELEGLQVKIQRLKLCRALOTE 461
DB 1255 RPKKTL-----BEPAKQLQEQ 1272

RESULT 48
Q922J3 PRELIMINARY; PRT; 1391 AA.
AC Q922J3;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Similar to restin (Reed-Steinberg cell-expressed intermediate
DE filament-associated protein).
GN RSN
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007191; AAH07191.1; -;
DR MGD; MGI:1928401; Rsn.
DR GO; GO:0005881; C:cytoplasmic microtubule; IDA.
DR GO; GO:0008017; F:microtubule binding; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR InterPro; IPR000938; CAP-Gly.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF01302; CAP_Gly; 2.
DR SMART; SM00343; Znf_C2HC; 1.
DR PROSITE; PS00845; CAP_Gly_1; 2.
DR PROSITE; PS02045; CAP_Gly_2; 2.
SQ SEQUENCE 1391 AA; 155813 MW; 061BED1FB3D4068D CRC64;

Query Match 9.0%; Score 244; DB 11; Length 1391;
Best Local Similarity 23.8%; Pred. No. 0.0008;
Matches 112; Conservative 86; Mismatches 153; Indels 120; Gaps 18;

QY 117 KEPSKGDPTNEIRQSEVGDGR-----DHRPQEKKAAGLGRK-----ITLIMQTLN 164
DB 734 KANSEKLELTLLQQLGAEKQIKNLETENAESSNANSITKELQEKELVLTGLQDSLN 793

QY 165 TLS-----TPBEKLAALCKVYALLEHRNSQOM-----KLLQKQ-----SOLVQEKD 209
DB 794 QNVQKTELEKELQTLKKEFASTSEBAVSAQTRMQDTVTKLHKEEQFNVLSELEKLE 853

QY 210 HLRGEHSK-----AVLARSKLSLCHRELQ---HNRS-----LKEEGVOR 246
DB 854 NLTDMAEKFEKDDREBQLVKAEKLENDIAEINKMGSDNSQLTKWDELRLKERSVEE 913

QY 247 AREBEKKEVTSHFOVTLNDIQLQMEQ-----HNERNKLRQENNELAERLKLIE 298

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:35:06 ; Search time 31 Seconds
(without alignments)
882.638 Million cell updates/sec

Title: US-10-023-529-8
Perfect score: 2702
Sequence: 1 KSSPQBPAGPEGAQRPSPQ.....APSTEASQGTGPQEPSTARA 530

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2702	100.0	530	4	US-08-979-608A-8
2	2702	100.0	530	4	US-09-517-849-8
3	2702	100.0	530	4	US-09-616-289-8
4	2693	99.7	546	4	US-09-616-289-44
5	2465.5	91.2	557	4	US-08-979-608A-5
6	2465.5	91.2	557	4	US-09-517-849-5
7	2465.5	91.2	557	4	US-09-616-289-5
8	1038.5	38.4	386	2	US-08-968-751-6
9	248.5	9.2	786	4	US-09-688-188B-23
10	248.5	9.2	786	4	US-09-291-417D-23
11	248.5	9.2	1001	4	US-09-688-188B-31
12	248.5	9.2	1001	4	US-09-291-417D-31
13	247	9.1	787	4	US-09-688-188B-151
14	247	9.1	787	4	US-09-291-417D-151
15	247	9.1	1001	3	US-09-060-410-2
16	247	9.1	1001	4	US-09-723-458-2
17	240	8.9	2662	4	US-09-595-684B-31
18	239.5	8.9	1898	1	US-08-056-200-94
19	239.5	8.9	1898	4	US-08-800-644-94
20	236.5	8.8	1898	4	US-09-310-187A-1
21	231.5	8.6	1886	4	US-08-938-105-3
22	230.5	8.5	3878	4	US-09-914-259-11
23	228	8.4	1388	2	US-08-685-576-1
24	227.5	8.4	1162	2	US-08-728-323A-2
25	227.5	8.4	1162	4	US-09-298-568-2
26	227.5	8.4	1162	4	US-09-410-395-2
27	226.5	8.4	3248	1	US-08-353-700-1

28	226.5	8.4	3248	5	PCT-US95-16216-1	Sequence 1, Appli
29	223	8.3	2101	1	US-08-195-487-4	Sequence 4, Appli
30	223	8.3	2101	5	PCT-US93-06160-4	Sequence 4, Appli
31	222	8.2	2101	1	US-08-466-390-4	Sequence 4, Appli
32	222	8.2	2101	1	US-08-470-950-4	Sequence 4, Appli
33	222	8.2	2101	1	US-08-467-781-4	Sequence 4, Appli
34	222	8.2	2101	2	US-08-483-924-4	Sequence 4, Appli
35	222	8.2	2101	3	US-09-452-294-1	Sequence 6, Appli
36	220.5	8.2	2482	1	US-08-328-254-6	Sequence 1, Appli
37	219.5	8.1	1695	4	US-09-866-108A-15753	Sequence 15753, A
38	218	8.1	443	2	US-08-795-475-6	Sequence 6, Appli
39	217.5	8.0	967	4	US-09-914-259-21	Sequence 21, Appli
40	217.5	8.0	1972	4	US-08-875-435B-4	Sequence 4, Appli
41	215.5	8.0	816	2	US-08-533-306A-6	Sequence 6, Appli
42	215.5	8.0	816	2	US-08-742-923A-6	Sequence 6, Appli
43	215.5	8.0	885	2	US-08-533-306A-4	Sequence 4, Appli
44	215.5	8.0	885	2	US-08-742-923A-4	Sequence 4, Appli
45	215.5	8.0	1354	3	US-08-685-871-2	Sequence 2, Appli
46	215	8.0	1388	4	US-09-976-594-296	Sequence 296, App
47	214	7.9	1388	2	US-08-685-576-4	Sequence 4, Appli
48	213	7.9	864	4	US-08-714-741-40	Sequence 40, Appli
49	212	7.8	2954	4	US-09-150-867-1	Sequence 1, Appli
50	211.5	7.8	1027	4	US-09-914-259-27	Sequence 27, Appli
51	211	7.8	1581	4	US-08-866-108A-15754	Sequence 15754, A
52	207	7.7	977	4	US-03-010-147B-18	Sequence 18, Appli
53	206.5	7.6	900	2	US-08-630-822A-62	Sequence 62, Appli
54	206.5	7.6	900	2	US-09-005-069-62	Sequence 62, Appli
55	205.5	7.6	900	4	US-08-171-156A-21	Sequence 21, Appli
56	206.5	7.6	900	4	US-09-004-730A-21	Sequence 21, Appli
57	206.5	7.6	900	4	US-08-981-799A-21	Sequence 21, Appli
58	206.5	7.6	1530	4	US-09-976-594-736	Sequence 736, App
59	206.5	7.6	1972	4	US-08-875-435B-3	Sequence 3, Appli
60	206	7.6	472	4	US-08-166-350-17	Sequence 17, Appli
61	206	7.6	611	4	US-09-216-393B-81	Sequence 81, Appli
62	206	7.6	1044	4	US-09-107-532A-2229	Sequence 5229, Ap
63	206	7.5	1180	4	US-09-543-681A-6436	Sequence 6436, Ap
64	203.5	7.5	976	3	US-09-104-324B-4	Sequence 4, Appli
65	203	7.5	2125	4	US-09-919-172-39	Sequence 29, Appli
66	202	7.5	8991	4	US-08-714-741-32	Sequence 32, Appli
67	201	7.4	748	4	US-08-688-188B-24	Sequence 24, Appli
68	201	7.4	748	4	US-09-291-417D-24	Sequence 24, Appli
69	200.5	7.4	534	4	US-09-103-664A-2	Sequence 2, Appli
70	200	7.4	1231	4	US-08-714-741-41	Sequence 41, Appli
71	199	7.4	458	4	US-08-529-055-73	Sequence 73, Appli
72	198.5	7.3	576	2	US-08-533-306A-2	Sequence 2, Appli
73	198.5	7.3	576	2	US-08-742-923A-2	Sequence 2, Appli
74	198.5	7.3	1032	4	US-09-914-259-26	Sequence 26, Appli
75	198	7.3	1279	4	US-09-724-517-2	Sequence 2, Appli
76	198	7.3	1279	4	US-09-641-807A-2	Sequence 2, Appli
77	198	7.3	1279	4	US-09-723-096-2	Sequence 2, Appli
78	197.5	7.3	1618	1	US-07-853-913-4	Sequence 4, Appli
79	197	7.3	993	3	US-03-050-410-4	Sequence 4, Appli
80	197	7.3	993	4	US-08-723-458-4	Sequence 4, Appli
81	196	7.3	966	4	US-08-688-188B-154	Sequence 154, App
82	196	7.3	966	4	US-09-291-417D-154	Sequence 154, App
83	195.5	7.2	898	4	US-09-688-188B-22	Sequence 22, Appli
84	195.5	7.2	898	4	US-09-291-417D-22	Sequence 22, Appli
85	195.5	7.2	1093	5	PCT-US93-03077-1	Sequence 1, Appli
86	193.5	7.2	627	4	US-09-702-953B-3	Sequence 3, Appli
87	193.5	7.2	647	4	US-09-702-953B-2	Sequence 2, Appli
88	192.5	7.1	716	4	US-09-219-983A-20	Sequence 20, Appli
89	192.5	7.1	1201	4	US-09-252-991A-32259	Sequence 32259, A
90	192	7.1	957	4	US-09-914-359-16	Sequence 16, Appli
91	192	7.1	1388	4	US-09-572-191-2	Sequence 2, Appli
92	192	7.1	1388	4	US-09-723-262-2	Sequence 2, Appli
93	192	7.1	1388	4	US-09-723-219-2	Sequence 2, Appli
94	191.5	7.1	456	4	US-09-808-701A-24	Sequence 24, Appli
95	191.5	7.1	803	4	US-09-154-750A-85	Sequence 85, Appli
96	191.5	7.1	803	4	US-09-665-479A-12	Sequence 12, Appli
97	191.5	7.1	1298	4	US-09-645-456A-14	Sequence 14, Appli
98	191.5	7.1	1298	4	US-09-425-324A-14	Sequence 14, Appli
99	191.5	7.1	1298	4	US-09-645-791-14	Sequence 14, Appli
100	191.5	7.1	1306	4	US-09-645-456A-10	Sequence 10, Appli

101 191.5 7.1 1306 4 US-09-425-324A-10 Sequence 10, Appl
102 191.5 7.1 1306 4 US-09-645-791-10 Sequence 10, Appl
103 191 7.1 961 4 US-09-914-259-66 Sequence 66, Appl
104 191 7.1 968 4 US-09-688-188B-107 Sequence 107, Appl
105 191 7.1 968 4 US-09-688-188B-155 Sequence 155, Appl
106 191 7.1 968 4 US-09-291-417D-107 Sequence 107, Appl
107 191 7.1 968 4 US-09-291-417D-107 Sequence 155, Appl
108 191 7.1 2310 4 US-09-874-923-120 Sequence 120, Appl
109 191 7.1 2568 4 US-09-866-108A-3 Sequence 3, Appl
110 190.5 7.1 973 4 US-09-392-714-24 Sequence 24, Appl
111 190 7.0 595 4 US-09-462-951B-1 Sequence 1, Appl
112 190 7.0 779 4 US-10-164-595-56 Sequence 56, Appl
113 190 7.0 911 4 US-09-688-188B-92 Sequence 92, Appl
114 190 7.0 911 4 US-09-291-417D-92 Sequence 92, Appl
115 189.5 7.0 963 4 US-09-914-259-20 Sequence 20, Appl
116 189.5 7.0 1164 4 US-09-457-708-2 Sequence 2, Appl
117 189.5 7.0 1164 4 US-09-950-046A-2 Sequence 2, Appl
118 189.5 7.0 1164 4 US-09-976-594-989 Sequence 989, Appl
119 188 7.0 956 4 US-09-914-259-17 Sequence 17, Appl
120 187.5 6.9 963 4 US-09-914-259-22 Sequence 22, Appl
121 187 6.9 574 4 US-09-702-953B-4 Sequence 4, Appl
122 186 6.9 467 2 US-08-686-599A-17 Sequence 17, Appl
123 186 6.9 912 4 US-09-688-188B-26 Sequence 26, Appl
124 186 6.9 912 4 US-09-291-417D-26 Sequence 26, Appl
125 185.5 6.9 982 4 US-09-688-188B-90 Sequence 90, Appl
126 185.5 6.9 982 4 US-09-291-417D-90 Sequence 90, Appl
127 185 6.8 432 2 US-08-686-599A-18 Sequence 18, Appl
128 185 6.8 433 2 US-08-686-599A-5 Sequence 5, Appl
129 185 6.8 493 2 US-08-686-599A-16 Sequence 16, Appl
130 185 6.8 588 4 US-08-714-741-42 Sequence 42, Appl
131 185 6.8 683 6 5210183-3 Patent No. 5210183
132 184.5 6.8 959 4 US-09-914-259-67 Sequence 67, Appl
133 184.5 6.8 1031 4 US-09-914-259-24 Sequence 24, Appl
134 184.5 6.8 1312 2 US-08-692-126-148 Sequence 148, Appl
135 184.5 6.8 1312 2 US-08-687-080-51 Sequence 51, Appl
136 184.5 6.8 1312 4 US-09-168-595-148 Sequence 148, Appl
137 184 6.8 591 4 US-09-252-891A-24904 Sequence 24904, A
138 183.5 6.8 704 4 US-09-370-838-191 Sequence 191, Appl
139 183.5 6.8 735 4 US-10-164-595-80 Sequence 80, Appl
140 183.5 6.8 784 4 US-10-164-595-79 Sequence 79, Appl
141 183.5 6.8 843 4 US-10-164-595-54 Sequence 54, Appl
142 183 6.8 1196 4 US-09-107-332A-3944 Sequence 3944, A
143 182 6.7 975 4 US-09-914-259-19 Sequence 19, Appl
144 182 6.7 1375 4 US-09-722-139-2 Sequence 2, Appl
145 182 6.7 1375 4 US-09-721-832-2 Sequence 2, Appl
146 182 6.7 1375 4 US-09-721-689-2 Sequence 2, Appl
147 181.5 6.7 1326 4 US-09-688-188B-15 Sequence 15, Appl
148 181.5 6.7 1326 4 US-09-291-417D-15 Sequence 15, Appl
149 181.5 6.7 1805 1 US-07-853-913-2 Sequence 2, Appl
150 181 6.7 468 4 US-09-328-352-6321 Sequence 6321, Ap

ALIGNMENTS

RESULT 1
US-08-979-608A-8
; Sequence 8, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Iaw, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA

COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/979,608A
FILING DATE: 26-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/048,547
FILING DATE: 03-JUN-1997
APPLICATION NUMBER: US 60/031,930
FILING DATE: 27-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-979-608A-8

Query Match 100.0%; Score 2702; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.2e-175;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSSQPEAGPEGAQPSQAAPVAREPGSSQAAPRPGQAQTAQSGALRDVSEELS 60
DB 1 KSSQPEAGPEGAQPSQAAPVAREPGSSQAAPRPGQAQTAQSGALRDVSEELS 60
QY 61 RQLEIDILSTYCVDDNQGGGEGDGAQEPAPEDAKSRITYVARNGEPPTPVVYGEKPS 120
DB 61 RQLEIDILSTYCVDDNQGGGEGDGAQEPAPEDAKSRITYVARNGEPPTPVVYGEKPS 120
QY 121 KGDPTTEIRQDEVDGDRHRRPQKKAKGLGKEITLLMQTLNTLSTPEEKLAALCKY 180
DB 121 KGDPTTEIRQDEVDGDRHRRPQKKAKGLGKEITLLMQTLNTLSTPEEKLAALCKY 180
QY 181 AELLBEHRNSQOMKLLQKQSQQLVQKDHLAGHSHKAVLARSKLESCLREIQRHNSLK 240
DB 181 AELLBEHRNSQOMKLLQKQSQQLVQKDHLAGHSHKAVLARSKLESCLREIQRHNSLK 240
QY 241 REGVQAREEERKREKVTSHFOVTLNDIQLQMEQHNRNSKLQRQNMELAEKLLIEQY 300
DB 241 REGVQAREEERKREKVTSHFOVTLNDIQLQMEQHNRNSKLQRQNMELAEKLLIEQY 300
QY 301 EUREEHIDKVFHKDLQQLVDKILQQAQEMLKEAEERHOREKDFLLKEAVSQRMCML 360
DB 301 EUREEHIDKVFHKDLQQLVDKILQQAQEMLKEAEERHOREKDFLLKEAVSQRMCML 360
QY 361 KQOETHLKQALALYTEKEEPQNTLSKSEVFTTFKQEMEKMTKKIKLEKETTYRSRW 420
DB 361 KQOETHLKQALALYTEKEEPQNTLSKSEVFTTFKQEMEKMTKKIKLEKETTYRSRW 420
QY 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLEKLCALQTERNDLNKRVDLSAGQGSLL 480
DB 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLEKLCALQTERNDLNKRVDLSAGQGSLL 480
QY 481 TDSGPERPEGGAQAPSPRVTEAPCYPGAPSTASGQTGPQEPPTSARA 530
DB 481 TDSGPERPEGGAQAPSPRVTEAPCYPGAPSTASGQTGPQEPPTSARA 530

RESULT 2

QY 361 KQETHLKQALALYTERFEFQNTLSKSSSEVFTTFKQEMERWTKIKKLEKETTYMRSW 420
Db 361 KQETHLKQALALYTERFEFQNTLSKSSSEVFTTFKQEMERWTKIKKLEKETTYMRSW 420
QY 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQSL 480
Db 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQSL 480
QY 481 TDSGPERFEGGQAQAPSSPRVTEAPCYPCAPSTASGQTGPQEPSARA 530
Db 481 TDSGPERFEGGQAQAPSSPRVTEAPCYPCAPSTASGQTGPQEPSARA 530

RESULT 4

US-09-616-289-44
; Sequence 44, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR FILING DATE: 2000-03-02
; PRIOR FILING DATE: 1997-11-26
; PRIOR FILING DATE: 1996-11-27
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-616-289-44

Query Match 99.7%; Score 2693; DB 4; Length 546;
Best Local Similarity 99.8%; Pred. No. 9.3e-175;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPGEAQPSPQAAPAVAEAGPGSSQAPKPEGAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPGEAQPSPQAAPAVAEAGPGSSQAPKPEGAQARTAQSGALRDVSEELS 76
QY 61 RQLEDILSTYCVDNQGGPGEAGPAPPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 120
Db 77 RQLEDILSTYCVDNQGGPGEAGPAPPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 136
QY 121 KGPNTTEIRQSDVGDHRRPQEKKAKGLGKITLLMOTLNTLSTPEEKLAALCKKY 180
Db 137 KGPNTTEIRQSDVGDHRRPQEKKAKGLGKITLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 AELLEERHSKQMKLLQKQSQVLQVQKHLRGEHSKAVLARSKLSLCRELQHRNRSK 240
Db 197 AELLEERHSKQMKLLQKQSQVLQVQKHLRGEHSKAVLARSKLSLCRELQHRNRSK 256
QY 241 EGVQARBEERKXEVTSFQVTLNDIQLOMEQHNRNRSKLRQENNELAERLAKLIEQY 300
Db 257 EGVQARBEERKXEVTSFQVTLNDIQLOMEQHNRNRSKLRQENNELAERLAKLIEQY 316
QY 301 ELREEHIDKVFHKDIQQQLVDKQLQQAQEMLKEAERHOREKDFLLKKAESQRMCELM 360
Db 317 ELREEHIDKVFHKDIQQQLVDKQLQQAQEMLKEAERHOREKDFLLKKAESQRMCELM 376
QY 361 KQETHLKQALALYTERFEFQNTLSKSSSEVFTTFKQEMERWTKIKKLEKETTYMRSW 420

Db 377 KQETHLKQALALYTERFEFQNTLSKSSSEVFTTFKQEMERWTKIKKLEKETTYMRSW 436
QY 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQSL 480
Db 437 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQSL 496
QY 481 TDSGPERFEGGQAQAPSSPRVTEAPCYPCAPSTASGQTGPQEPSARA 530
Db 497 TDSGPERFEGGQAQAPSSPRVTEAPCYPCAPSTASGQTGPQEPSARA 546

RESULT 5

US-08-979-608A-5
; Sequence 5, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; TITLE OF INVENTION: BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TITLE OF INVENTION: TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/08/979,608A
; FILING DATE: 26-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-08-979-608A-5

Query Match 91.2%; Score 2465.5; DB 4; Length 557;
Best Local Similarity 90.8%; Pred. No. 2.5e-159;
Matches 492; Conservative 12; Mismatches 25; Indels 13; Gaps 3;

QY 1 KSSPGQPEAGPGEAQPSPQAAPAVAEAGPGSSQAPKPEGAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPGEAQPSPQAAPAVAEAGPGSSQAPKPEGAQARTAQSGALRDVSEELS 75
QY 61 RQLEDILSTYCVDNQGGPGEAGPAPPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 119
Db 76 RQLEDILSTYCVDNQGGPGEAGPAPPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 135

;/ NUMBER OF SEQ ID NOS: 155
;/ SOFTWARE: PatentIn Ver. 2.1
;/ SEQ ID NO 31
;/ LENGTH: 1001
;/ TYPE: PRT
;/ ORGANISM: Artificial Sequence
;/ FEATURE:
;/ OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;/ OTHER INFORMATION: murine/human SULU3
US-09-688-188B-31

Query Match 9.2%; Score 248.5; DB 4; Length 1001;
Best Local Similarity 21.6%; Pred. No. 8.6e-09;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

QY	23	PAVEAE-----GPGSSQA-PRKPEGAQARTAQSGALRDVSEELSRQLE	64
Db	324	PAVEAQEEBEEQDHGVRGTGVNSVGSNQSIPTSMISASSQSSVNSLFDVSDOKS-ELD	382
QY	65	DILSTCYVNNQO-----GPGEDG--AQGPAPEDAEKSRITYVARNGEPPTPVVYGEKE	118
Db	383	MMEGDHTVMSNSSVHLKPEENYREGDP-----RT---RASDPQSPQVSRHKS	430
QY	119	PSKGDPTNTEIROSDEVDGRDHRPQEKKAAGLGEIT-----LLMQTLNTLSTPEEKLA	174
Db	431	HYNRREHPATIRTASLV---TQMQEHEQDSLEQMSGYKMRROHQKQMLTLENKLK	486
QY	175	ALCKKYA-----ELLEHRN--SOKMKLLQKQSOLVQE-----KDHLCGEHKS	217
Db	487	AMDEHRLDKOLETORNFAAEKEKLTKHQAAEKEAKVMSNEEKKFQOHIOAQQCK	546
QY	218	AVLARSKLESICRELQHRNRSKEGVQR-----AREEEE---KRKEVTSHFQV--TLNDI	268
Db	547	EL--NSFLESQKREYKLKKEELNENQSTPKKEQEWLSKQENIQHFQABEEANLL	604
QY	269	QLQWE-----QHN-----ENSKLRQENMELAEKLLKLEIQEYELRE	304
Db	605	RRQROYLELCRRFKRMLLGRNLEQDLVREELNKROTQKDLHAMLRLQHESMQELEF	664
QY	305	EHIDKVFKHDLQOOLVDALQQAQAEMLKEAEERHOREKDFLLKEAVESQRMCELMKQOE	364
Db	665	RHLNTIQK--MRCELL--RLQHQTELNTQLEYNKRERELRKHVMVEVRQPKSLKSKE	719
QY	365	THLKQOL-----ALYTEKPEFQNTLSKSEVFTTFQEMEKMTYKIKKLEKETMYRSRW	420
Db	720	LQIKKQFQDTCKIQTRYKALRNHLE-----TTPKSEHKAVLKRLK--BEQTRKLAILA	772
QY	421	ESSNKALLEMAEKTVR-----DKELE	442
Db	773	EQYDHSINEMLSQALRLDEAQAEQCVLKMQLQOELELLNAYQSKIKNQAEACHDELR	832
QY	443	GLQVKI-----QRLEKLCRALQTERND-----LNKRVDLSAGGQGS-----L	480
Db	833	ELQRVSLRRALLEQKIEEEMALQNERTERISLLERQARIEAFDESMLGFSNMVL	892
QY	481	TDSGPERPEGPAQAPSSPRVTEAPCYGAP--STEASQOTGP	522
Db	893	SNLSPE-----AFSHSYPGASGWSHNPTGGPGP	920

RESULT 13
US-09-688-188B-151
;/ Sequence 151 Application US/09688188B
;/ Patent No. 656716
;/ GENERAL INFORMATION:
;/ APPLICANT: PLOWMAN, GREGORY
;/ APPLICANT: MARTINEZ, RICARDO
;/ APPLICANT: WHYTE, DAVID
;/ TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
;/ FILE REFERENCE: 038602/0328
;/ CURRENT APPLICATION NUMBER: US/09/688.188B
;/ CURRENT FILING DATE: 2000-10-16

;/ NUMBER OF SEQ ID NOS: 155
;/ SOFTWARE: PatentIn Ver. 2.1
;/ SEQ ID NO 31
;/ LENGTH: 1001
;/ TYPE: PRT
;/ ORGANISM: Artificial Sequence
;/ FEATURE:
;/ OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;/ OTHER INFORMATION: murine/human SULU3
US-09-688-188B-31

Query Match 9.2%; Score 248.5; DB 4; Length 1001;
Best Local Similarity 21.6%; Pred. No. 8.6e-09;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

QY	23	PAVEAE-----GPGSSQA-PRKPEGAQARTAQSGALRDVSEELSRQLE	64
Db	324	PAVEAQEEBEEQDHGVRGTGVNSVGSNQSIPTSMISASSQSSVNSLFDVSDOKS-ELD	382
QY	65	DILSTCYVNNQO-----GPGEDG--AQGPAPEDAEKSRITYVARNGEPPTPVVYGEKE	118
Db	383	MMEGDHTVMSNSSVHLKPEENYREGDP-----RT---RASDPQSPQVSRHKS	430
QY	119	PSKGDPTNTEIROSDEVDGRDHRPQEKKAAGLGEIT-----LLMQTLNTLSTPEEKLA	174
Db	431	HYNRREHPATIRTASLV---TQMQEHEQDSLEQMSGYKMRROHQKQMLTLENKLK	486
QY	175	ALCKKYA-----ELLEHRN--SOKMKLLQKQSOLVQE-----KDHLCGEHKS	217
Db	487	AMDEHRLDKOLETORNFAAEKEKLTKHQAAEKEAKVMSNEEKKFQOHIOAQQCK	546
QY	218	AVLARSKLESICRELQHRNRSKEGVQR-----AREEEE---KRKEVTSHFQV--TLNDI	268
Db	547	EL--NSFLESQKREYKLKKEELNENQSTPKKEQEWLSKQENIQHFQABEEANLL	604
QY	269	QLQWE-----QHN-----ENSKLRQENMELAEKLLKLEIQEYELRE	304
Db	605	RRQROYLELCRRFKRMLLGRNLEQDLVREELNKROTQKDLHAMLRLQHESMQELEF	664
QY	305	EHIDKVFKHDLQOOLVDALQQAQAEMLKEAEERHOREKDFLLKEAVESQRMCELMKQOE	364
Db	665	RHLNTIQK--MRCELL--RLQHQTELNTQLEYNKRERELRKHVMVEVRQPKSLKSKE	719
QY	365	THLKQOL-----ALYTEKPEFQNTLSKSEVFTTFQEMEKMTYKIKKLEKETMYRSRW	420
Db	720	LQIKKQFQDTCKIQTRYKALRNHLE-----TTPKSEHKAVLKRLK--BEQTRKLAILA	772
QY	421	ESSNKALLEMAEKTVR-----DKELE	442
Db	773	EQYDHSINEMLSQALRLDEAQAEQCVLKMQLQOELELLNAYQSKIKNQAEACHDELR	832
QY	443	GLQVKI-----QRLEKLCRALQTERND-----LNKRVDLSAGGQGS-----L	480
Db	833	ELQRVSLRRALLEQKIEEEMALQNERTERISLLERQARIEAFDESMLGFSNMVL	892
QY	481	TDSGPERPEGPAQAPSSPRVTEAPCYGAP--STEASQOTGP	522
Db	893	SNLSPE-----AFSHSYPGASGWSHNPTGGPGP	920

RESULT 12
US-09-291-417D-31
;/ Sequence 31 Application US/09291417D
;/ Patent No. 668170
;/ GENERAL INFORMATION:
;/ APPLICANT: PLOWMAN, GREGORY
;/ APPLICANT: MARTINEZ, RICARDO
;/ APPLICANT: WHYTE, DAVID
;/ TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
;/ FILE REFERENCE: 038602/0329
;/ CURRENT APPLICATION NUMBER: US/09/291.417D
;/ CURRENT FILING DATE: 1999-04-13
;/ PRIOR APPLICATION NUMBER: 60/081,784

; PRIOR APPLICATION NUMBER: 09/291,417
; PRIOR FILING DATE: 1998-04-14
; PRIOR APPLICATION NUMBER: 60/081,784
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 151
; LENGTH: 787
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-688-188B-151

Query Match
Best Local Similarity 9.1%; Score 247; DB 4; Length 787;
Matches 139; Conservative 117; Mismatches 206; Indels 178; Gaps 29;

QY 14 AQERPSQAAPAVEAGP-----GSSQA-PRKPEGAQAQTAQSGALRDVSEELSRQLEDILS 68
DB 114 AQEEBEEQDHGVRGTGTVNSVGSNQSIPMSISASSQSSVNSLPSVDDSKS-ELDMWEG 172
QY 69 TYCVNNQG---GPGEDG--AQCEPAEPEDAESRTTVARNGEPEPTPVYVGEKPSKG 122
DB 173 DHTVNSNSVHLKPEENYREGDP-----RT---RASDPQSPQVSRHKSRYRN 220
QY 123 DPNTTEIRQSDVGDHRRPQEKKAKGLGKEIT---LLMQTLNTLSTPEEKLAALCK 178
DB 221 REHFATIRTASLV---TRQMEHEQDSSELREQMSGYKMRERQKQKQTLNKLKAEWD 276
QY 179 KYA-----ELLEHRN--SOKOMKLLQKQSOVLQV-----KDLRGEHSHKAVIA 221
DB 277 EHRLLDKCLSTGRNFAAEMKLIKQQAAMEKAKVMSNEEKFKQOHIQAQKKEL-- 334
QY 222 RSKLESCLRELQHRNLSKEGVOR---AREEBE---KRKEVTSHFQV--TLNDIQLQM 272
DB 335 NSFLESQKEEYKLRKEQLKEELNENQSTPKKEQEWLSKOKENIQHFOABEENLLRRQR 394
QY 273 E-----OHN-----ERNSKLRQENMELAEKLIKLEQYELREEHID 308
DB 395 QYLELECRFRKRMMLGRHNLQDLVBEELNKROTQKDLHAMLROHESMCELEFRHLN 454
QY 309 KYFKHKLQOQLVDKLAQQAQEMLEKEAREHOREKDFLLKEAVESQRMCELMKQOETHLK 368
DB 455 TIQK---MRCELI--RLQHOTELTNQLEYNKRERELRRKHVMEVRQPKSLKSKELQIK 509
QY 369 QQL---ALYTEKPEFQNTLSKSEVFTTFKQEMEMTKIKKLEKETMYRGRWESSN 424
DB 510 KQFQDTCKIQTRYKALRNHLE---TTPKSEHKAVALKRLK--EQOTRKLAILAEQYD 562
QY 425 KALLEMAEKTVR-----DKELEGLOV 446
DB 563 HSNEXMLSTQALRLDRAQAEACQVLKMLQOQLELLNAYQSKIKWQAQAQHDRLRELEQ 622
QY 447 KI-----QRLEKLCRALQTERND-----LNKRVQDLSAGGGS-----LTDG 484
DB 623 RVSLRRALLEQKIEEEMALQNERTERIRSLRQAREIEAFDSESMLGFSNNVLSNLS 682
QY 485 PRRPEGGAQAPSSPRVTPACVPGAP--STEASGOTGP 522
DB 683 PE-----AFSHSYFGASGWSHNPTGGFGP 706

RESULT 14
US-09-291-417D-151
; Sequence 151, Application US/09291417D
; Patent No. 6880170
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0329
; CURRENT APPLICATION NUMBER: US/09/291,417D
; CURRENT FILING DATE: 1999-04-13

; PRIOR APPLICATION NUMBER: 60/081,784
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 151
; LENGTH: 787
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-291-417D-151

Query Match
Best Local Similarity 9.1%; Score 247; DB 4; Length 787;
Matches 139; Conservative 117; Mismatches 206; Indels 178; Gaps 29;

QY 14 AQERPSQAAPAVEAGP-----GSSQA-PRKPEGAQAQTAQSGALRDVSEELSRQLEDILS 68
DB 114 AQEEBEEQDHGVRGTGTVNSVGSNQSIPMSISASSQSSVNSLPSVDDSKS-ELDMWEG 172
QY 69 TYCVNNQG---GPGEDG--AQCEPAEPEDAESRTTVARNGEPEPTPVYVGEKPSKG 122
DB 173 DHTVNSNSVHLKPEENYREGDP-----RT---RASDPQSPQVSRHKSRYRN 220
QY 123 DPNTTEIRQSDVGDHRRPQEKKAKGLGKEIT---LLMQTLNTLSTPEEKLAALCK 178
DB 221 REHFATIRTASLV---TRQMEHEQDSSELREQMSGYKMRERQKQKQTLNKLKAEWD 276
QY 179 KYA-----ELLEHRN--SOKOMKLLQKQSOVLQV-----KDLRGEHSHKAVIA 221
DB 277 EHRLLDKCLSTGRNFAAEMKLIKQQAAMEKAKVMSNEEKFKQOHIQAQKKEL-- 334
QY 222 RSKLESCLRELQHRNLSKEGVOR---AREEBE---KRKEVTSHFQV--TLNDIQLQM 272
DB 335 NSFLESQKEEYKLRKEQLKEELNENQSTPKKEQEWLSKOKENIQHFOABEENLLRRQR 394
QY 273 E-----OHN-----ERNSKLRQENMELAEKLIKLEQYELREEHID 308
DB 395 QYLELECRFRKRMMLGRHNLQDLVBEELNKROTQKDLHAMLROHESMCELEFRHLN 454
QY 309 KYFKHKLQOQLVDKLAQQAQEMLEKEAREHOREKDFLLKEAVESQRMCELMKQOETHLK 368
DB 455 TIQK---MRCELI--RLQHOTELTNQLEYNKRERELRRKHVMEVRQPKSLKSKELQIK 509
QY 369 QQL---ALYTEKPEFQNTLSKSEVFTTFKQEMEMTKIKKLEKETMYRGRWESSN 424
DB 510 KQFQDTCKIQTRYKALRNHLE---TTPKSEHKAVALKRLK--EQOTRKLAILAEQYD 562
QY 425 KALLEMAEKTVR-----DKELEGLOV 446
DB 563 HSNEXMLSTQALRLDRAQAEACQVLKMLQOQLELLNAYQSKIKWQAQAQHDRLRELEQ 622
QY 447 KI-----QRLEKLCRALQTERND-----LNKRVQDLSAGGGS-----LTDG 484
DB 623 RVSLRRALLEQKIEEEMALQNERTERIRSLRQAREIEAFDSESMLGFSNNVLSNLS 682
QY 485 PRRPEGGAQAPSSPRVTPACVPGAP--STEASGOTGP 522
DB 683 PE-----AFSHSYFGASGWSHNPTGGFGP 706

RESULT 15
US-09-060-410-2
; Sequence 2, Application US/09060410
; Patent No. 6155461
; GENERAL INFORMATION:
; APPLICANT: Cobb, Melanie
; APPLICANT: Hutchinson, Michele
; APPLICANT: Chen, Zhu
; APPLICANT: Bertram, Kevin
; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/060,410
FILING DATE: 14-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.421
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1001 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-060-410-2

Query Match 9.1%; Score 247; DB 3; Length 1001;
Best Local Similarity 21.9%; Pred. No. 1.1e-08;
Matches 142; Conservative 118; Mismatches 212; Indels 176; Gaps 30;

QY 23 PAVEAE-----GPGSSQA-PRKPGAQAARTAGSALRDVSELSRQLE 64
DB 324 PAVEAESEEBQDHGGRTGTWNSVGNQSIIPMSISASSQSSVNSLPDASDDKS-ELD 382
QY 65 DILSTYCVNNQ-----GPGEDGAQGEPAEPDAEKSRITYVARNGEPEPTPVVYGEKPS 120
DB 383 MMEGDTWMSNSVHLKPEENYQ-EEGDR-----TRASAPQSPQVSRHSHY 432
QY 121 KGPNTTEIRQSDVEGDRHRRPORKKAGLGKEIT-----LLMQTLNTLSTPEKLAAL 176
DB 433 RNEHFATIRTSALV-----TRQMEHQDSBELRQMSGYKMRQRQHQKQLMTLENKLA 488
QY 177 CKKYA-----ELLEHRN--SOKQMLLQKQSOVLQV-----KDLRGEHSHY 219
DB 489 MDEHRLDKOLETQRNFAAEKLIKQASMEKEAKYMANEKKFOOHIAQOQKEL 548
QY 220 LARSKLESCLRELQRNHSKKEGVOR-----ARBEES--KXKVTSHFQV--TLNDIQL 270
DB 549 --NSPLESQKRYKLRKBQKEELNENQSTPKKEKQEWLSKQENIQHFQAEFEANLLR 606
QY 271 QMS-----QHN-----ERNSKLRQNMELABRLKXLIQYELREH 306
DB 607 QRYLELECRFRKRMMLGHNLLEQDLYRELNKQTKDLEHMLLRHSHESMQLFEX 666
QY 307 IDKVPKHQLOQOLVDKALQQAQEMLKAEERHQKOPFLKEAVESQRMCLMQQETH 366
DB 667 LNTIQK---MRCELI--RLQEQTELNTQLEYKRRERELRRKHVMVEVRQPKSLSKELQ 721
QY 367 LKQOL-----ALYTEKEFEETNTLSKSEVFTTPKQEMKMTKKIKLEKFTMYRSWES 422
DB 722 IRKQFOOTCKIQTRQYKAUNHLE-----TTPKSEHKAVLKKL--ESQTRKLAIAEQ 774
QY 423 SNKALLEMAEKTVR-----DKELEGL 444
DB 775 YDHSINEMLSLQALRLDEAQAEQVLMQLOQOELELLNAYQSKIWKQAEQAQHDRELRL 834
QY 445 QVKI-----QRLKLCALQTERND-----LNRVQDLSGGQGS-----LTD 482
DB 835 EQRYSRRALLEQKIEEMALQNERTHIRSLRQAEIEAFQDSMRLGFSNNVLN 894

QY 483 SGPERRPEG-PAQAQS-SPRVTRAP--CYFGAPSTEASG---QTGPGQ 523
DB 895 LSPFAFSHYFCASSWSHNPCTGCGSPHMGHDMGCTPOAWGHPMQGGPQ 942
RESULT 15
US-09-723-458-2
Sequence 2, Application US/09723458
Patent No. 6586242
GENERAL INFORMATION:
APPLICANT: Cobb, Melanie
Hutchinson, Michele
Chen, Zhu
Berman, Kevin
TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEND AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/723,458
FILING DATE: 27-NO. 6586242-2000
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/060,410
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.421
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1001 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-723-458-2

Query Match 9.1%; Score 247; DB 4; Length 1001;
Best Local Similarity 21.9%; Pred. No. 1.1e-08;
Matches 142; Conservative 118; Mismatches 212; Indels 176; Gaps 30;
QY 23 PAVEAE-----GPGSSQA-PRKPGAQAARTAGSALRDVSELSRQLE 64
DB 324 PAVEAESEEBQDHGGRTGTWNSVGNQSIIPMSISASSQSSVNSLPDASDDKS-ELD 382
QY 65 DILSTYCVNNQ-----GPGEDGAQGEPAEPDAEKSRITYVARNGEPEPTPVVYGEKPS 120
DB 383 MMEGDTWMSNSVHLKPEENYQ-EEGDR-----TRASAPQSPQVSRHSHY 432
QY 121 KGPNTTEIRQSDVEGDRHRRPORKKAGLGKEIT-----LLMQTLNTLSTPEKLAAL 176
DB 433 RNEHFATIRTSALV-----TRQMEHQDSBELRQMSGYKMRQRQHQKQLMTLENKLA 488
QY 177 CKKYA-----ELLEHRN--SOKQMLLQKQSOVLQV-----KDLRGEHSHY 219
DB 489 MDEHRLDKOLETQRNFAAEKLIKQASMEKEAKYMANEKKFOOHIAQOQKEL 548
QY 220 LARSKLESCLRELQRNHSKKEGVOR-----ARBEES--KXKVTSHFQV--TLNDIQL 270

Db 549 --NSFLESOKREYKLRKQELNENOSTPKKQKQEMLSKQENIQHPQAEERANLRR 606
QY 271 QME-----QHN-----ERNKLRQENMELAEKLLIQEYELREBH 306
Db 607 QOYLECECRFPKRMMLGRHNLQDLYVEELNKRTQKDLHMLRLQHESQELERFH 666
QY 307 IDKVFHKDLOOLVDKIQOQKQEMLEAEERHOREKOPFLKEAVESQKCELMQOETH 366
Db 667 LNTIQK---MRCELI---RLQHOTELTNQLEYKRRERELRRHVMVEVROQPKSLAKELQ 721
QY 367 LKQOL---ALVTEKPEEPONTLSKSSSEVFTTFKQEMERMTKKIKLEKETTMYRSWES 422
Db 722 IKKQODTKCIQTRQYKALRNHLZ-----TTPKSEHKAVLRKX--SEQTKLAILAEQ 774
QY 423 SNKALLEMAEBKTVR-----DKELEGE 444
Db 775 YDHSINEMLSQALRLDEAEACQVLKMQLOQLELLNAYQSKIMQWAEAHQDRELREL 834
QY 445 QVKI-----ORLEKLCRALQTERND-----LNKRVODLSAGCGS-----LTD 482
Db 835 EQRVSRLRALLQKIEEMALQNERIRISLLERQAREIFAQSESRVLGFSNMVLN 894
QY 483 SOPERPEG-PGAQAPS-SPRVTEAP--CYPGAPSTEASG---QTGPQ 523
Db 895 LSPFAPSHSYPGASSMHNPTGSGPHMGHPMGGTPQAWGHPMGQGPQ 942

RESULT 17

US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766

; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe

; APPLICANT: Ohashi, Cara

; APPLICANT: Sakowicz, Roman

; APPLICANT: Vaiberg, Eugene

; APPLICANT: Wood, Kenneth

; APPLICANT: Yu, Ming

; TITLE OF INVENTION: Human kinesins and methods of producing

; FILE REFERENCE: cytop036

; CURRENT APPLICATION NUMBER: US/09/595,684B

; PRIOR FILING DATE: 2002-06-24

; PRIOR APPLICATION NUMBER: 09/295,612

; NUMBER OF SEQ ID NOS: 105

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 31

; LENGTH: 2662

; TYPE: PRT

; ORGANISM: Human

US-09-595-684B-31

Query Match 8.9%; Score 240; DB 4; Length 2662;

Best Local Similarity 20.2%; Pred. No. 1,1e-07;

Matches 130; Conservative 101; Mismatches 176; Indels 238; Gaps 25;

QY 55 VSEHLSROLEDILSTYCDNNQGGPGEAGQ-----EPAPPEDAKSTYVA----- 102

Db 986 ISEEVSRNLH-----MEENTGETKDFQQRWGWGDKKQDLKAKNTQTLTADVKDNEII 1038

QY 103 -----RNGEPPTPVYCEKPSKGD-----PNTTEIR-QSDEVGDRDH 140

Db 1039 EQRKIFSLIOBKNELOQWLSVIAEKQDLTKENIEMTIENQSELRLLGDLKKQOE 1098

QY 141 RFPQEKKA-KGLGKEITLLMTLSTPBEKL----- 173

Db 1099 IVAQEKNAIKKEGE-----LSRTCDRLAEVEEKLKESQQLQEQQLNQEEMSEMQ 1154

QY 174 -----AALCKYAEELLEHNRSOKMKLQKQKQSLVQS 207

Db 1155 KINEIENLKNELKNKELTLEHMETERLELAQKLENYEVKSIYKRVKVLQKSPETE 1214

QY 174 -----AALCKYAEELLEHNRSOKMKLQKQKQSLVQS 207

Db 1155 KINEIENLKNELKNKELTLEHMETERLELAQKLENYEVKSIYKRVKVLQKSPETE 1214

QY 174 -----AALCKYAEELLEHNRSOKMKLQKQKQSLVQS 207

Db 1155 KINEIENLKNELKNKELTLEHMETERLELAQKLENYEVKSIYKRVKVLQKSPETE 1214

QY 174 -----AALCKYAEELLEHNRSOKMKLQKQKQSLVQS 207

Db 1155 KINEIENLKNELKNKELTLEHMETERLELAQKLENYEVKSIYKRVKVLQKSPETE 1214

QY 174 -----AALCKYAEELLEHNRSOKMKLQKQKQSLVQS 207

Db 1155 KINEIENLKNELKNKELTLEHMETERLELAQKLENYEVKSIYKRVKVLQKSPETE 1214

QY 174 -----AALCKYAEELLEHNRSOKMKLQKQKQSLVQS 207

Db 1155 KINEIENLKNELKNKELTLEHMETERLELAQKLENYEVKSIYKRVKVLQKSPETE 1214

QY 174 -----AALCKYAEELLEHNRSOKMKLQKQKQSLVQS 207

Db 1155 KINEIENLKNELKNKELTLEHMETERLELAQKLENYEVKSIYKRVKVLQKSPETE 1214

QY 174 -----AALCKYAEELLEHNRSOKMKLQKQKQSLVQS 207

Db 1155 KINEIENLKNELKNKELTLEHMETERLELAQKLENYEVKSIYKRVKVLQKSPETE 1214

QY 208 KDHLRG-----EHSKAV--LARSKLESCL-----RELQRHN 236
Db 1215 RDHLRGYRRETEATGLQTKELKIAHILKEHQETIDELRSVSEKTAQIINTQDLBKSH 1274
QY 237 RSLKEEGVQVAREEEREKEV--TSHFOVTLNDIOLQMEQHNEHS-----KLROENMELA 290
Db 1275 TGLQEE-IPVLEHQQELLPNVKYVSETOETNNLELLTEQSTTKDSTTLARIEMERLRN 1333
QY 291 ERLKKLIQYELREBHIDKV-----FKHKLQOOLVD--AKLQQAQ----- 329
Db 1334 EKFBQSQBEIKSLTKERDNLTKIKEALEVKGHDQLKEHIRETLAKIQESQSQEQSLANKKE 1393
QY 330 -----EML---KEAERHOREBKDFLLKEAVESQKCELMQOETH 359
Db 1394 KONETTKIVSEMOFKPKDSALLRIETIEMLGSLKLOESHDEMKS-VAKEXDDQLQLOEV 1452
QY 360 MKQETHLKKQOL---ALYTEKPEEFO-----NTLSKSESEVFTTFKQ 397
Db 1453 LQSSDQLKENIKVAKHLSLEELKVAHCLKEQSEINELRVNLSEKETEISTIQKQ 1512
QY 398 ---EMEKTKIKKL-EKETM---YRSWESSKALLEMAEKKTVRDKELEG-QVKI-- 448
Db 1513 LEANDKLQNKIQBIYEKEBQANIKOISEVOENVELKQFKEHRAKQDSALQSIESKMLE 1572
QY 449 -----ORLEKLCRALQTERNDLNKRVODLSA 474
Db 1573 LTNRLQSQBEIQIMIXEKEMKRVQELQIERDQLKENTYKIVA 1617

RESULT 18

US-08-056-200-94

; Sequence 94, Application US/08056200

; Patent No. 5616500

; GENERAL INFORMATION:

; APPLICANT: Steinert, Peter M.

; APPLICANT: Lee, Seung-Chul

; APPLICANT: Kim, In-Gyu

; APPLICANT: Chung, Soo-Il

; APPLICANT: Park, Sang-Chul

; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and

; TITLE OF INVENTION: Methods of Using Same

; NUMBER OF SEQUENCES: 117

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe, Martens, Olson & Bear

; STREET: 620 Newport Center Drive, Sixteenth Floor

; CITY: Newport Beach

; STATE: CA

; COUNTRY: U.S.A.

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA: US/08/056,200

; FILING DATE: 30-APR-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Pedrick, Michael P.

; REGISTRATION NUMBER: 36,799

; REFERENCE/DOCKET NUMBER: NIH054.001A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (714) 760-0404

; TELEFAX: (714) 760-9502

; INFORMATION FOR SEQ ID NO: 94:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1898 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-056-200-94

; TITLE OF INVENTION: Graft Rejection
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310,187A
; CURRENT FILING DATE: 1999-05-12
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-310-187A-1

Query Match 8.8%; Score 236.5; DB 4; Length 1939;
Best Local Similarity 24.4%; Pred. No. 1.3e-07;
Matches 146; Conservative 81; Mismatches 196; Indels 175; Gaps 25;

Qy 12 EGACERPSQAAPVAEABGPGSSQAPKPEGAQARTAGSALRDVSELSQLEDILSTYC 71
Db 1037 EGSLEQEKVMDLE-----RAKRLKEDLKLQES--IMDLND-KLQLEKLLKKE 1086

Qy 72 VDNNGGPGDGAQGEPAPEDA-----EKSRTYVAR-----NGPEPTPVVYGEKEP 119
Db 1087 FDIHQ-----QNSKIEDEQALALQLOKLEKQARIEEELBELEAERTAKVLEKL 1138

Qy 120 SKGDPNTEIROS-DEVG-----DRHRRPQKKKAKGLGKBITLLMOTLNTLSTPEE 171
Db 1139 SDLSRELEISERLEAGGATSQVIEENKKGRAEFQKWRDLLENTL-----QHEA 1189

Qy 172 KLAALCKKY-----ABLLLEHRNSQOMKLLQKQKOSQOLVQKDHRLRGHSHKAVLARSKLS 227
Db 1190 TAAALRKHADSVAEIGEIQIDNLRVKQKLEKSEFPLELDVTSNMEOIIKAKANLEK 1249

Qy 228 LCR-----ELQRNRSLEKBEGVORAR----- 248
Db 1250 VSRTELDQANRYVKLEEAQRSLNDFTTORAKLOTENGELARQLEKEALISQLTGKLS 1309

Qy 249 -----ESEEKKEVTSH-FQVTLNDIQLOMEQHNER-----NSX 281
Db 1310 YTOQMEDLKQLEBEGKAKNALAHALQSAHDCDLLREQYEEETEAKAELQRLVLSKANSE 1369

Qy 282 LR-----QENMELAEKLLIEQYELREEHIDKV-----FKHK---DLQQ 318
Db 1370 VAQWRTKYETDAIQRTTELEBAKKLQALQLODAEEAVEAVNAKCSSLEKTKHRLQWEIED 1429

Qy 319 QLVDAKLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLQKQALALYTEKF 378
Db 1430 LMVDVERSNA---AAALDKKQNFDKILAE--WKQKYESSQSELESSQKARSLSLTLP 1484

Qy 379 EEFONTLSKSGSEVFTTFKQEMKMTKKI-----KKLEKETMYR 417
Db 1485 -KLKAYEESLEHLETFKRNKNLOEISDLTEQLGEGGKNNVHELEKVRKQLEVEKLELQ 1543

Qy 418 SRWESSNKALLEMAEERTVDRKLEGLQVKIQRLEKLC---RALQTERNDLAKRVQD 471
Db 1544 SALEEA-EASLEHEEGKILR-AQLSFNQIKABIERKLAKEDEMEQAKRM--HQRVVD 1597

RESULT 21
US-08-938-105-3
; Sequence 3, Application US/08938105
; Patent No. 6353151
; GENERAL INFORMATION:
; APPLICANT: Weinwand, Leslie A.
; APPLICANT: Vikstrom, Karen L.
; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,105
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Crook, Wamell M.
REGISTRATION NUMBER: 31,071
REFERENCE/DOCKET NUMBER: 3595-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1886 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-938-105-3

Query Match 8.6%; Score 231.5; DB 4; Length 1886;
Best Local Similarity 24.4%; Pred. No. 2.7e-07;
Matches 146; Conservative 80; Mismatches 197; Indels 175; Gaps 25;

Qy 12 EGACERPSQAAPVAEABGPGSSQAPKPEGAQARTAGSALRDVSELSQLEDILSTYC 71
Db 984 EGSLEQEKVMDLE-----RAKRLKEDLKLQES--IMDLND-KLQLEKLLKKE 1033

Qy 72 VDNNGGPGDGAQGEPAPEDA-----EKSRTYVAR-----NGPEPTPVVYGEKEP 119
Db 1034 FDIHQ-----QNSKIEDEQALALQLOKLEKQARIEEELBELEAERTAKVLEKL 1085

Qy 120 SKGDPNTEIROS-DEVG-----DRHRRPQKKKAKGLGKBITLLMOTLNTLSTPEE 171
Db 1086 SDLSRELEISERLEAGGATSQVIEENKKGRAEFQKWRDLLENTL-----QHEA 1136

Qy 172 KLAALCKKY-----ABLLLEHRNSQOMKLLQKQKOSQOLVQKDHRLRGHSHKAVLARSKLS 227
Db 1137 TAAALRKHADSVAEIGEIQIDNLRVKQKLEKSEFPLELDVTSNMEOIIKAKANLEK 1196

Qy 228 LCR-----ELQRNRSLEKBEGVORAR----- 248
Db 1197 VSRTELDQANRYVKLEEAQRSLNDFTTORAKLOTENGELARQLEKEALISQLTGKLS 1256

Qy 249 -----ESEEKKEVTSH-FQVTLNDIQLOMEQHNER-----NSX 281
Db 1257 YTOQMEDLKQLEBEGKAKNALAHALQSAHDCDLLREQYEEEMEAKAELQRLVLSKANSE 1316

Qy 282 LR-----QENMELAEKLLIEQYELREEHIDKV-----FKHK---DLQQ 318
Db 1317 VAQWRTKYETDAIQRTTELEBAKKLQALQLODAEEAVEAVNAKCSSLEKTKHRLQWEIED 1376

Qy 319 QLVDAKLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLQKQALALYTEKF 378
Db 1377 LMVDVERSNA---AAALDKKQNFDKILAE--WKQKYESSQSELESSQKARSLSLTLP 1431

Qy 379 EEFONTLSKSGSEVFTTFKQEMKMTKKI-----KKLEKETMYR 417
Db 1432 -KLKAYEESLEHLETFKRNKNLOEISDLTEQLGEGGKNNVHELEKVRKQLEVEKLELQ 1490

Qy 418 SRWESSNKALLEMAEERTVDRKLEGLQVKIQRLEKLC---RALQTERNDLAKRVQD 471
Db 1491 SALEEA-EASLEHEEGKILR-AQLSFNQIKABIERKLAKEDEMEQAKRM--HQRVVD 1544

RESULT 22
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336

GENERAL INFORMATION:
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-685-576-1

Query Match 8.4%; Score 228; DB 2; Length 1388;
Best Local Similarity 21.0%; Pred. No. 3.2e-07;
Matches 130; Conservative 83; Mismatches 177; Indels 228; Gaps 20;

QY 91 PDAEKS-----TYVARGEP-----PTPW--- 113
DB 316 PDAEKSRAKALICAFITDREVRIGRVEIKOHPFFQNDQWWDNIRETAAVPVPL 375
QY 114 -----YGEKPSKGDPTTEE-----ROSDEVG 136
DB 376 SSDIDSSNFDDIDDDGDVETFPPIKAFVGNQLPFIFTYYRENLILLSPSCKENDSIQ 435
QY 137 DRDHRPQ-----KKAKGLGKEITLLMQTL 163
DB 436 SRNEESQIQKLYLLEHLSTEIOAKEELQCKSVNTRLEKVAKELEBEITLRKVE 495
QY 164 NTLSTPEKLAALCKYAE-----LLEHRNSOKMK 195
DB 496 STLRLQREKALLQHKNAEYQKADHEADKKNLENDVNSLKQLEDLKKRNQNSQISTE 555
QY 196 LLOKQSQVLQVKDHLRGEHSKAVLAR-----SKLELCRELORHNSLKEEGVQ 245
DB 556 KYNQLQRQDQETNALLRTESDTAARLKTQAEBSKQIQOLESNNRDLQKNCLELTAKLK 615
QY 245 RARE-----EERKKEVTSHFQVTLNDIQ-----LQMEQHN 276
DB 616 LEKEPILQSVLESEERDR--THGSEIINDLQGRISGLEEDVANGKILLAKVELEKRLQ 673

GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2006-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-914-259-11

Query Match 8.5%; Score 230.5; DB 4; Length 3878;
Best Local Similarity 18.4%; Pred. No. 7.7e-07;
Matches 119; Conservative 110; Mismatches 195; Indels 223; Gaps 23;

QY 31 GSSQAPRPGAQARTASGALRVSELSRQLDILSTYCVNNO-----GGPGEHG 83
DB 161 GAQSPHLEMMSELA--GRQHEI--BELARELEEMVTVTGEGLQQLQSFRAAIKORDG 217
QY 84 -----AQGEPAEPEDAERSYVARGPEPTPVVYGEKPSKGDPTNT----- 126
DB 218 IITOLTANLQAREKEDTWREFLELTQSQKLIQFQQLQASETLNTHSSTAADLLQ 277
QY 127 ---EIRQSDEVDGRDHRPQEKKANGKGLKRTILMTLNTLSTPEKLAALCKK----- 179
DB 278 AKQIILTHQQLQEDQLLEDYQKKEDFTVQISFLQEKIKVYEMEQDKVENSKEEIQ 337
QY 180 -----YAELEHRNSOKM-----KILQKQSOLVQKDLHGEHSKAVLAR 223
DB 338 EKETIIEELNTKIIEEKKTILELKDXTADKLLGELQELQVQKQEI--NMKLEITNS 395
QY 224 K-----LESCLRELQNR--SLKEGVQARAREBEKRYE----- 256
DB 396 KQKERSSEBIKQLMGTVLEQKXNHDQSQFETDIVQRMQETQKLEQLRAELDEMYG 455
QY 257 -----VTSH-----FOYTLNDIOLQME 273
DB 456 QIVQMKQELTROHMAQHEEMKTRKGENALRSYNTVNEQIKLMTVAINELIKLO 515
QY 274 QHNRNSKLK-----QENNELAERLKLTEQVELREHIDKVK-----HK 314
DB 516 DTNSQKELKEELGLILBEKCALQRLQEDLVEELSFREIQIARQTIARQESKLINEAK 575
QY 315 DLQO-QLVADKLOAQEMLKEABERH-----OREKDFLLKEAVESQRM--- 356
DB 576 SLSTVEDLKAIVSASSRKELELKHAEVTVNYKIKLEMLEKEKNAVDNRMAESQEALE 635
QY 357 -----CELMKQOE-----TLKQQL-----ALYTEKF 378
DB 636 RLRTQLLFSHEBELSKLEDLEIHRINIEBKLNGLIHYKQIDGLQNMESQKIETMQF 695
QY 379 BEFQNTLSKSEVF-----TTFKQEMKQTKIKKLEKETTMYRSRNSKAL 427
DB 696 EK-DNLITKQNLILEISKLDQOOSLVNSKSEMTLOINELQKEIEILR----- 744
QY 428 LEMAEKTVRDLQEGQVKIQRLKCLRALQTERNDLNRKRVQDLA 474
DB 745 -QEEKKGTLQEQVQELQKTELLEK---QMKKENDLQKFAOLEA 787

RESULT 23
US-08-685-576-1
; Sequence 1, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Kozi
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Nakano, Takeshi

QY 172 KLAALCKVYALLESERNSQKMLKQKQSQVQKDKHLRGHSHKAVLARSKLESICRE 231
DB 641 QDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDEQ 700
QY 232 LQRHNSLKEGVQARER 291
DB 701 EQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDE 760
QY 292 RLKLLIEQVELREHIDKVFHKDLQQLVDAKLQQAQEMLKAEERHOREKDFLLKBAV 351
DB 761 QQEQ 818
QY 352 ESQRMELMKQEQTHLKKQALYTERFEFQNTLSKSSVFTTFKQEMKMTKKIKKLEK 411
DB 819 EEQEQ-ELSEEQ 875
QY 412 ETTMYRSWESSNKALLEAEKTVRDKELGLOVKIQRLKCLRALOTERNDLNKRVQD 471
DB 876 QEQEQELEVEEQ 931
QY 472 -----LSAGGQGSULTDGPERRPFGCAQAPSPRVTEAPCPGAPSTASQGTGP 522
DB 932 TVEPIILHGSSSEDEMEVDYVVTSTHQIASSP-----PGDNTFDDDDPQGP 979
RESULT 26
US-09-410-399-2
; Sequence 2, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; FILE REFERENCES: UM-03778
; CURRENT APPLICATION NUMBER: US/09/410,399
; FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-2

Query Match 8.4%; Score 227.5; DB 4; Length 1162;
Best Local Similarity 19.4%; Pred. No. 2.7e-07;
Matches 104; Conservative 128; Mismatches 261; Indels 43; Gaps 13;
QY 1 KSSPGQPEAGPEGAQRPSSQAAPAVAEFGPGSQAPKPEGAQARTAGALRDVSEELS 60
DB 473 QDEQ 528
QY 61 ROLEDILSTYCVNNGGPGEDGAGPEP--ABEDAKSRITYVARNGEPP--BPTPVVYG 115
DB 529 PQQOE-----PQQOEPPQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ 580
QY 116 EKPSKDPNTERIRQDE----VGDRHRRPQKKGKGLGKEITLLMQLTNTLSTPER 171
DB 581 QREPPQREPPQREPPQREPPQREPPQREPPQREPPQREPPQREPPQREPPQREPPQ 640
QY 172 KLAALCKVYALLESERNSQKMLKQKQSQVQKDKHLRGHSHKAVLARSKLESICRE 231
DB 641 QDEQ 700
QY 232 LQRHNSLKEGVQARER 291
DB 701 EQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDEQEQDE 760
QY 292 RLKLLIEQVELREHIDKVFHKDLQQLVDAKLQQAQEMLKAEERHOREKDFLLKBAV 351
DB 761 QQEQ 818

QY 352 ESQRMELMKQEQTHLKKQALYTERFEFQNTLSKSSVFTTFKQEMKMTKKIKKLEK 411
DB 819 EEQEQ-ELSEEQ 875
QY 412 ETTMYRSWESSNKALLEAEKTVRDKELGLOVKIQRLKCLRALOTERNDLNKRVQD 471
DB 876 QEQEQELEVEEQ 931
QY 472 -----LSAGGQGSULTDGPERRPFGCAQAPSPRVTEAPCPGAPSTASQGTGP 522
DB 932 TVEPIILHGSSSEDEMEVDYVVTSTHQIASSP-----PGDNTFDDDDPQGP 979
RESULT 27
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
US-08-353-700-1

Query Match 8.4%; Score 226.5; DB 1; Length 3248;
Best Local Similarity 22.4%; Pred. No. 1.2e-06;
Matches 102; Conservative 91; Mismatches 191; Indels 71; Gaps 14;
QY 31 GSSQAPRPEGAQARTAGALRDVSEELSROLEDILSTYCVNNGGPGEDGAGPEAP 90
DB 272 GNSSPHLLDQK--AQOELRNKINLELFL-----QCH--E 305
QY 91 PEDAKSRITYVARNGEPPPTPVYGEKPSKDPNTERIRQSDVGDHRRPQKCKAK 150
DB 306 KSKQGVNKFQELQLEKAVKVELIEKVLNKCDELVRITTAQYDQASTKTALQK 365
QY 151 GLGKEITLLMQLTNTL-STPEKLAALCKVYALLESERNSQKMLKQK-----QSQV 205

Db 366 KLTDLSQORNAESARCSELEQKIK--EKEKPEQBELSRQORSFQTLDDQECIQMKARLT 422
QY 206 QSKDHLRGHSAVLARSKLSLSCRELQHRNLSKEGVQVORABEEKKEVTSHFQVTL 265
Db 423 QELQAKMHNVLQALDKLTSVKQLENNLEEFKQ--KLCRAEQAFQAS-----QIKE 474
QY 266 NDIQLOMEQHNERNKLRQENNELA-----ERLKKLIEQYELREBHIDKVPKHKDLOQO 319
Db 475 NEILRSMEEMKKNLLKSHSEQKAREVCHLEAEKLNKIQCLNOSQNFPAEEMKAKNTSOE 534
QY 320 LVDAKLQQAQEMKAEERHQRKDFLLKAEAVESQRMCM--ELMKQOETHLKQQLALYTEK 377
Db 535 TM---LRDLQKINQOENSLTLEKLAVALDEKQDCSDLLKKREHHI-----581
QY 378 FREFQNTLSKSEVFTTFKQEMEKMTKKIKLEKETTMYRSWESSNKALLEMABEKTVR 437
Db 582 -EQLNDKLSKTEKESKALLSALSELKCKEYBELKEETLF-SCKWSENEKLL-----TQM 633
QY 438 DKELEGLOVKIQLEKLCRALQTERNDKRVQDL 472
Db 634 ESEKENLQSKINHLETCLTKQIQKSHYNERVRTL 668
RESULT 28
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dana, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1
Query Match 8.4%; Score 226.5; DB 5; Length 3248;
Best Local Similarity 22.4%; Pred. No. 1.2e-06;
Matches 102; Conservative 91; Mismatches 191; Indels 71; Gaps 14;
QY 31 GSSQAPKPEGAQARTAQSGALRDVSELSRQLEDILSTVCVNNQGGPGEDCAQCEPAB 90

Db 272 GNSSPHLLDQK---AQNLKRNKINELEURL-----QGH--E 305
QY 91 PEDAEKSRITYVARNGEPEPTPVVYGEKPSKGDPTBEIRQSDVEGDRHRRPOEKKKAK 150
Db 306 KEMGQVNFQELQLEKAKVLELIEKVKLNKCRDELVRTTAQYDQASTKYTALEQKXK 365
QY 151 GLGKEITLMTLNTL-STPEEKALALCKKVAELLEHRNSQKQKLLQKK-----QSQIV 205
Db 366 KLTDLSQORNAESARCSELEQKIK--EKEKPEQBELSRQORSFQTLDDQECIQMKARLT 422
QY 206 QSKDHLRGHSAVLARSKLSLSCRELQHRNLSKEGVQVORABEEKKEVTSHFQVTL 265
Db 423 QELQAKMHNVLQALDKLTSVKQLENNLEEFKQ--KLCRAEQAFQAS-----QIKE 474
QY 266 NDIQLOMEQHNERNKLRQENNELA-----ERLKKLIEQYELREBHIDKVPKHKDLOQO 319
Db 475 NEILRSMEEMKKNLLKSHSEQKAREVCHLEAEKLNKIQCLNOSQNFPAEEMKAKNTSOE 534
QY 320 LVDAKLQQAQEMKAEERHQRKDFLLKAEAVESQRMCM--ELMKQOETHLKQQLALYTEK 377
Db 535 TM---LRDLQKINQOENSLTLEKLAVALDEKQDCSDLLKKREHHI-----581
QY 378 FREFQNTLSKSEVFTTFKQEMEKMTKKIKLEKETTMYRSWESSNKALLEMABEKTVR 437
Db 582 -EQLNDKLSKTEKESKALLSALSELKCKEYBELKEETLF-SCKWSENEKLL-----TQM 633
QY 438 DKELEGLOVKIQLEKLCRALQTERNDKRVQDL 472
Db 634 ESEKENLQSKINHLETCLTKQIQKSHYNERVRTL 668
RESULT 29
US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOURKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,487
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,701
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-195-487-4

Query Match 8.3%; Score 223; DB 1; Length 2101;
Best Local Similarity 22.1%; Pred. No. 1.2e-06;
Matches 111; Conservative 91; Mismatches 229; Indels 72; Gaps 17;

QY 42 AQARTAGSALRDVSELSQLEIDILSTYCVNNQGGPCED-GAOGEPAPEDAEKSRVY 100
DB 1296 ASGREAEK--QRVASENLRLQEL-----TSQAEAEELGQELKAWQKEFFQKEQAL 1343
QY 101 VARGEPPTPVVYGEKPSKGDPTNTEIRQSDVEGDRDHR--POEKKVAKGLGKEITL 158
DB 1344 STLQLEHTSTQALVSELLPAK--HLCCQQAQAEAAKREHRELEQSKQAAGGLRAELLR 1401
QY 159 LMQTLNLTSTPEKLAALCKYAELEHRSNOKMLKQKOSQOLVQEKDHLRGHESKA 218
DB 1402 AQELGELIPLRQKVAEQDRTAQQLRAEKASVYAEQLSMLKKAHGLLAEN---RGLGERA 1458
QY 219 VLARSKLESLEL--QRNRSLSKEGVQVQARAEEREEKREKVTSHFQVTLNDIOLQMBHNE 277
DB 1459 NLGRQFLV---ELDQAREKYVQELAAVRA--DAETRLAEVQREBAQSTARELEVMTAKYEG 1514
QY 278 RNSKLRENMELEAKLLEQVLEEREHIDKVFHKDLQOQLVDKLAQQAQEMLKEABE 337
DB 1515 AKVLEERQRFQERQKLTQAV-----ELSKKLADS-----DOASK 1552
QY 338 RHQREKDFLLKAVESQRCMELKQOEHLKQOLALYTEKPEEFQNTLSKSEVFTTFKQ 397
DB 1553 VQOQKLKAVQAQGGESQQAQRFQALNELQALSKQAEHYKLMQEKAKTHYDAKKQ 1612
QY 398 EMEKMTKIKKLEKETTMYRSWESSNKALLEWABEKTVRDKELEGLQVKIQLEKLCRA 457
DB 1613 QNQELOQSLSE-----QLOKENKELRAEAE---LGHLEQAGLTKAEQTCRH 1661
QY 458 LOTERNDLNKR---QDLGAGGQSL--TDSGPERPEGPGCAQAPSSPRVTEAPCPYGA- 511
DB 1662 LTAQVRSLEAQVAHADQQLRDLGKFQVATDALKSRFPQAK-PQLDLSIDSLDLSCEBGP 1720

QY 512 -----PSTEASQGTGPQEPS 527
DB 1721 LSITSKLPRTPQDGTSPGEPAS 1743

RESULT 30
PCT-US93-06160-4
Sequence 4, Application PC/TUS9306160
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA HURWITZ & THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06160
FILING DATE: 19930621
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/248-7000
TELEFAX: 617/248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-06160-4

Query Match 8.3%; Score 223; DB 5; Length 2101;
Best Local Similarity 22.1%; Pred. No. 1.2e-06;
Matches 111; Conservative 91; Mismatches 229; Indels 72; Gaps 17;

QY 42 AQARTAGSALRDVSELSQLEIDILSTYCVNNQGGPCED-GAOGEPAPEDAEKSRVY 100
DB 1296 ASGREAEK--QRVASENLRLQEL-----TSQAEAEELGQELKAWQKEFFQKEQAL 1343
QY 101 VARGEPPTPVVYGEKPSKGDPTNTEIRQSDVEGDRDHR--POEKKVAKGLGKEITL 158
DB 1344 STLQLEHTSTQALVSELLPAK--HLCCQQAQAEAAKREHRELEQSKQAAGGLRAELLR 1401
QY 159 LMQTLNLTSTPEKLAALCKYAELEHRSNOKMLKQKOSQOLVQEKDHLRGHESKA 218
DB 1402 AQELGELIPLRQKVAEQDRTAQQLRAEKASVYAEQLSMLKKAHGLLAEN---RGLGERA 1458
QY 219 VLARSKLESLEL--QRNRSLSKEGVQVQARAEEREEKREKVTSHFQVTLNDIOLQMBHNE 277
DB 1459 NLGRQFLV---ELDQAREKYVQELAAVRA--DAETRLAEVQREBAQSTARELEVMTAKYEG 1514
QY 278 RNSKLRENMELEAKLLEQVLEEREHIDKVFHKDLQOQLVDKLAQQAQEMLKEABE 337
DB 1515 AKVLEERQRFQERQKLTQAV-----ELSKKLADS-----DOASK 1552
QY 338 RHQREKDFLLKAVESQRCMELKQOEHLKQOLALYTEKPEEFQNTLSKSEVFTTFKQ 397
DB 1553 VQOQKLKAVQAQGGESQQAQRFQALNELQALSKQAEHYKLMQEKAKTHYDAKKQ 1612
QY 398 EMEKMTKIKKLEKETTMYRSWESSNKALLEWABEKTVRDKELEGLQVKIQLEKLCRA 457
DB 1613 QNQELOQSLSE-----QLOKENKELRAEAE---LGHLEQAGLTKAEQTCRH 1661
QY 458 LOTERNDLNKR---QDLGAGGQSL--TDSGPERPEGPGCAQAPSSPRVTEAPCPYGA- 511
DB 1662 LTAQVRSLEAQVAHADQQLRDLGKFQVATDALKSRFPQAK-PQLDLSIDSLDLSCEBGP 1720

QY 512 -----PSTEASQGTGPQEPS 527
DB 1721 LSITSKLPRTPQDGTSPGEPAS 1743

RESULT 31
US-08-466-390-4
Sequence 4, Application US/08466390
Patent No. 5686562
GENERAL INFORMATION:
APPLICANT: TOUTATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,950
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELEPHONE: (617) 248-7100
TELEFAX: (617) 248-7000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-390-4

Query Match 8.2%; Score 222; DB 1; Length 2101;
Best Local Similarity 22.1%; Pred. No. 1.4e-06;
Matches 111; Conservative 91; Mismatches 229; Indels 72; Gaps 17;
QY 42 AQAATAGALRDVSEELSRQLEDILSTYCVNNQGGPDED-GAAGEPAEPDPAKSRTY 100
DB 1296 ASGREAEK--QRVASENLRLQEL-----TSQAEAEELQELKAWQKFFQKEQAL 1343
QY 101 VARNGEPEPTPVVYGEKPSKGDPTNTEIRQSDVGGDRHR--POEKKAAGLGEITL 158
DB 1344 STLQLEHTSTQALVSELLPAK--HLCQQLQAEQAAAEKREHELESGQQAAGLRAELLR 1401
QY 159 LMOTLNTLSTPEEKLAALCKKYAELEHNSOKMKLLQKQSQOLVQKQHLRGEHSA 218
DB 1402 AQELGELIPLRQVABQERTAQQLRAEKASYAEQLSMLKKAHGLAEEN---RGLGERA 1458
QY 219 VLARSKLESICREL-QRHNRSLEKEGVQARAEERKEEKEVTSHFQVTLNDIQLOMEQHNE 277
DB 1459 NEGROFLEV---ELDQAREKYVQELAAVRA--DAETRLAEVQREASTARELEVMTAKYEG 1514
QY 278 RNSKLRQENMELAEELKXLIQYELREHEHDKVFKHDKLQQLVDAKLOQAQEMLKEABE 337
DB 1515 AKVLEERQRFQERQKLTQAVE-----ELSKKLADS-----DQASK 1552
QY 338 RHOREKDFLLKEAVESQRMCELMKQOETHLKOOLALYTEKEFEFQNTLSKSSSVFTTFKQ 397
DB 1553 VQOQKLVQVQAGGSGEQAEQAFQALNELQALSQKEQAAEHYKLOMEKAKTHYDAKQ 1612
QY 398 EMERKTKIKKLEKTTMYRSRWESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRA 457
DB 1613 QNQLQELQSLRLE-----QLQENKELAEABR---LGHLEQAGLKTKEAQTCTRH 1661
QY 458 LQTERNDLAKRV----QDLISAGGQSL--TDSGPERPPEGPGCAAPSSPRVTEAPCPGA- 511
DB 1662 LTAQVRSLEAQVAHADQQLDLKGFQVATDALKSREPOAK-PQLDLSIDSLDSCBEGTP 1720
QY 512 -----PSTEASGQTGPQEPPTS 527
DB 1721 LSITSKLPRTPQDGTSPVGEFAS 1743

RESULT 32
US-08-470-950-4
Sequence 4, Application US/08/470,950
Patent No. 5698439
GENERAL INFORMATION:
APPLICANT: TOUKATLY, GARY
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSER: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET

CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,950
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELEPHONE: (617) 248-7100
TELEFAX: (617) 248-7000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-950-4

Query Match 8.2%; Score 222; DB 1; Length 2101;
Best Local Similarity 22.1%; Pred. No. 1.4e-06;
Matches 111; Conservative 91; Mismatches 229; Indels 72; Gaps 17;
QY 42 AQAATAGALRDVSEELSRQLEDILSTYCVNNQGGPDED-GAAGEPAEPDPAKSRTY 100
DB 1296 ASGREAEK--QRVASENLRLQEL-----TSQAEAEELQELKAWQKFFQKEQAL 1343
QY 101 VARNGEPEPTPVVYGEKPSKGDPTNTEIRQSDVGGDRHR--POEKKAAGLGEITL 158
DB 1344 STLQLEHTSTQALVSELLPAK--HLCQQLQAEQAAAEKREHELESGQQAAGLRAELLR 1401
QY 159 LMOTLNTLSTPEEKLAALCKKYAELEHNSOKMKLLQKQSQOLVQKQHLRGEHSA 218
DB 1402 AQELGELIPLRQVABQERTAQQLRAEKASYAEQLSMLKKAHGLAEEN---RGLGERA 1458
QY 219 VLARSKLESICREL-QRHNRSLEKEGVQARAEERKEEKEVTSHFQVTLNDIQLOMEQHNE 277
DB 1459 NEGROFLEV---ELDQAREKYVQELAAVRA--DAETRLAEVQREASTARELEVMTAKYEG 1514
QY 278 RNSKLRQENMELAEELKXLIQYELREHEHDKVFKHDKLQQLVDAKLOQAQEMLKEABE 337
DB 1515 AKVLEERQRFQERQKLTQAVE-----ELSKKLADS-----DQASK 1552
QY 338 RHOREKDFLLKEAVESQRMCELMKQOETHLKOOLALYTEKEFEFQNTLSKSSSVFTTFKQ 397
DB 1553 VQOQKLVQVQAGGSGEQAEQAFQALNELQALSQKEQAAEHYKLOMEKAKTHYDAKQ 1612
QY 398 EMERKTKIKKLEKTTMYRSRWESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRA 457
DB 1613 QNQLQELQSLRLE-----QLQENKELAEABR---LGHLEQAGLKTKEAQTCTRH 1661
QY 458 LQTERNDLAKRV----QDLISAGGQSL--TDSGPERPPEGPGCAAPSSPRVTEAPCPGA- 511
DB 1662 LTAQVRSLEAQVAHADQQLDLKGFQVATDALKSREPOAK-PQLDLSIDSLDSCBEGTP 1720
QY 512 -----PSTEASGQTGPQEPPTS 527
DB 1721 LSITSKLPRTPQDGTSPVGEFAS 1743

RESULT 33
US-08-467-781-4
Sequence 4, Application US/08/467,781
Patent No. 5780596

GENERAL INFORMATION:
APPLICANT: TOURKATLY, GARY
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,781
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-467-781-4

Query Match 8.2%; Score 222; DB 1; Length 2101;
Best Local Similarity 22.1%; Pred. No. 1.4e-06;
Matches 111; Conservative 91; Mismatches 229; Indels 72; Gaps 17;

QY 42 AQARTAGSALRDVSEELSRQLEDILSTYCVNNQGGPGED-GAAGEPAEPEDAERSRTY 100
DB 1296 ASGREAEK--QRVASENLRQEL-----TSQARAEEELGQELKAWQKFFQEQAL 1343
QY 101 VARGEPEPTPVVYGEKPSKGDPTNTEIRQSDVEGDRDHR--POEKKKAGLGKEITL 158
DB 1344 STLQLEHTSTQALVSELLPAK--HLCQQLAQEAQAAEKKRHEBELQSKQAAGLRAELLR 1401
QY 159 LMQTLNTLSTPEEKLAALCKKYAELEHNRNSQKMKLQKQSQLVQEKDHLRGEHSA 218
DB 1402 AQRELGLIPLRQKVAEQRTAQQLRAEKASVAEQLSMLKKAHGLLAEN---RGLGERA 1458
QY 219 VLARSKLESICREL--QRHNRSLKEEGVQRAEERBEKKREVTSHFOVTLNDIQLQMEQNE 277
DB 1459 NLGROFLV---ELDQAREKYVQELAAVRA--DAETRLAEVQRAQSTARELEVMTAKYEG 1514
QY 278 RNSKLRQENMELARLKKLIEQYELREHIDKVFKHDLQOQLVDAKLQQAQEMLKEABE 337
DB 1515 AKVVLVEERQRFQERQKTAQVE-----ELSKKLADS-----DQASK 1552
QY 338 RHOREKDFLLKEAVESQRMCHLMQOETHLQOQLALYTEKEEPONTLSKSEVFTTQK 397
DB 1613 QNGELOQLASLE-----QLQKENKELRAEAER---LGHLOQAGLKTKEAEOCTCRH 1661
QY 458 LQTERNDLNRV-----ODLSAGCGSL--TDSGPERBEPGGAQAPSPVTRAPCPGA- 511
DB 1662 LTAQVRSLEAQAHAHQDLRGLKQVATDALKSREPQAK--POLDUSIDLSLDCSEGT 1720

QY 512 -----PSTASGQTGPQBPTS 527
DB 1721 LSITSKLPRTQPGTSPVGPAS 1743

RESULT 34

US-08-483-924-4
Sequence 4, Application US/08483924
Patent No. 5882876
GENERAL INFORMATION:
APPLICANT: TOURKATLY, GARY
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-483-924-4

Query Match 8.2%; Score 222; DB 2; Length 2101;
Best Local Similarity 22.1%; Pred. No. 1.4e-06;
Matches 111; Conservative 91; Mismatches 229; Indels 72; Gaps 17;

QY 42 AQARTAGSALRDVSEELSRQLEDILSTYCVNNQGGPGED-GAAGEPAEPEDAERSRTY 100
DB 1296 ASGREAEK--QRVASENLRQEL-----TSQARAEEELGQELKAWQKFFQEQAL 1343
QY 101 VARGEPEPTPVVYGEKPSKGDPTNTEIRQSDVEGDRDHR--POEKKKAGLGKEITL 158
DB 1344 STLQLEHTSTQALVSELLPAK--HLCQQLAQEAQAAEKKRHEBELQSKQAAGLRAELLR 1401
QY 159 LMQTLNTLSTPEEKLAALCKKYAELEHNRNSQKMKLQKQSQLVQEKDHLRGEHSA 218
DB 1402 AQRELGLIPLRQKVAEQRTAQQLRAEKASVAEQLSMLKKAHGLLAEN---RGLGERA 1458
QY 219 VLARSKLESICREL--QRHNRSLKEEGVQRAEERBEKKREVTSHFOVTLNDIQLQMEQNE 277
DB 1459 NLGROFLV---ELDQAREKYVQELAAVRA--DAETRLAEVQRAQSTARELEVMTAKYEG 1514
QY 278 RNSKLRQENMELARLKKLIEQYELREHIDKVFKHDLQOQLVDAKLQQAQEMLKEABE 337
DB 1515 AKVVLVEERQRFQERQKTAQVE-----ELSKKLADS-----DQASK 1552
QY 338 RHOREKDFLLKEAVESQRMCHLMQOETHLQOQLALYTEKEEPONTLSKSEVFTTQK 397
DB 1553 VQOQKLKAVQAQGGESQQAQRFQALNELQAQLSQKEQAABHYKLQMEKAKTHYDAKQ 1612

QY 341 -----REKDFLLKEAVESORM-CELMKQOETHLKQALALYTEKTEFBQNTL 395
DB 2083 LDTNKQYEVITYREKLTKEECLSSQKLEIDLKSSKEELANSKATTOHLEELKTK 2142
QY 386 SKSEVPTTFKQEMERKTKIKLEKETTWYRSRWSSNKAL--LEMAEK-----TVRDK 439
DB 2143 MDNLKYVNLKKNERRAQGRMKLLKSKCKQLEBEKEILOKELSOQAQEKQKTGVMGT 2202
QY 440 ELSEGLQVIR-----LEKLC-----RALOTERDNLKRVQDLSAGGQ 478
DB 2203 KVELDTTEIKELKETLEETKEADEYLDKYCSLLISHKLEKAKEMLETOVAHLCS--QQ 2260
QY 479 SLTDS--GPRRPEGCAQAPSS--PRVTEAPCYFGAPSTEASGO 519
DB 2261 SKQDSRGSLGPWFG---PSPIPSVTEKELSSG--QNKASCK 2299

RESULT 37
US-09-866-108A-15753
; Sequence 15753, Application US/09866108A
; Patent No. 6686188
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharon G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AROMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108A
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 15755
; SOFTWARE: Aromica Sequence Listing Engine
; Patent No. 6686188
; SEQ ID NO 15753
; LENGTH: 1695
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-866-108A-15753

Query Match 8.1%; Score 219.5; DB 4; Length 1695;
Best Local Similarity 21.8%; Pred. No. 1.5e-06;
Matches 123; Conservative 86; Mismatches 219; Indels 137; Gaps 20;
QY 19 SQAAPAVEAPGSSQAPRPEGAQARTAGSALRVSELSQLEBILSTVCVDNNOGG 78
DB 1169 TOKVSLAELOPISQESKDEASLAKVKQ--LRDLKAKVQDEBEL-----DEQAG 1219
QY 79 PGEDGAGEPABEDAEKSR-TYVARNGEPTFVVYGEKEPSKGPNTHEIRQ----- 131

DB 1220 SIOMLEQAKLRLEMENRQTH-----SKEMESRDEFEVEARQSCOKL 1264
QY 132 -----SDEVGDRD-----HRRPQEKKAAGLGKEI 156
DB 1265 KQMEVQLEBEYEDKQALREKRELESKLSLTSDQVNRQDFSEKLRDLKTKALLADA 1324
QY 157 TLLMTQTLNTLTPPEKLAALCKYAELEEHNSQKMKLQKQKQKQKQKQKQKQKQKQKQ 216
DB 1325 QIMDLKNNAPSKKEIAQLKQ-----LEE-----SEFTCAAAVAKAKMEVEEDLHLOID 1377
QY 217 KAVLARSKLESICRELQRNRSALKKEGVQVRAFEERKEV-----TSHFQVTLN 266
DB 1378 DIAKAKTALAEQLSRLQRE---KNEIQNRLEEDQDMXELMKKHAQAAVAQASRDMAQKN 1433
QY 267 DILOMEQHNSKLRQENMELAKLLEIYELREHIDKVFYKHKDLOOQALVDKQ 326
DB 1434 DLQAQIERSN-----KEQELQKQALQSQVEF-----LEQSMWDKSLV 1473
QY 327 QAQE--MLXAEERHOREKDFLKAEVQSORMCMLKQOETHLKQOALALYTEKTEFBQNTL 385
DB 1474 SRQEKIRLETRLEFTEKTQVKRLENLASRLKETWELTEERDQRAA--AENREKEQN-- 1529
QY 386 SKSSEVPTTFKQEMERKTKIKLEKETTWYRSRWSSNKALLEMAEETVYDKELEGLO 445
DB 1530 KRLQRLDFTKESEELARK---EABASRKKEHEMDLES--LEAANOSLOAD-----LK 1579
QY 446 VKIORLEKLCRALQYER-----NDLNKRVQDLSAGQGGSLT--DSGPERPE 490
DB 1580 LAFKIGDLOAIBDEMSEDNEDLINSEGDSDVDSLEDRVDGVKSWLSKNKGFSPKAPS 1639
QY 491 GPGAQAPSSPRVTEAPCYFGAPSTE 515
DB 1640 DDGSLKSSSPTSHWKPLAPDPDSDE 1664

RESULT 38
US-08-795-475-6
; Sequence 6, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bivrick, Lars
; APPLICANT: Sivbring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-795-475-6

Query Match 8.1%; Score 218; DB 2; Length 443;
Best Local Similarity 23.4%; Pred. No. 3.6e-07;
Matches 111; Conservative 86; Mismatches 178; Indels 100; Gaps 21;

QY 104 NCEPPTVYVYGEKSKDPNTEETR-----QDEVGDRDHRHPPQEKKAK 150
DB 1 NGDGNPREVI---EDLAANPAIOTRLRHENKDLKARLENAMVAVAGDFKRAHLEKAK 57
QY 151 GLGKEITLLMOTL-NTLSTPEEKLAALCKYKABLLHEHRNSQOM--KLQKKQSQLOV 207
DB 58 -----QALEQDKDLETYKELQDDY-DLAKESTWDRQRLKELEBEKKEKALEA 106
QY 208 KQHLRGEHSAVLAARSKLSLREEL-----ORHNR-SLAKGQVQARBEKEKEKVT 258
DB 107 IQASRDYHRAALEKELBEKKALELAIDQASQDYNRANVLEKELETTTRQETNRNL 166
QY 259 SHFQVTLNDIOLQMEQHNRNSKLQOE-----NMELAEHLKLLIQ--YELR 303
DB 167 GNAKLELDLSSEKEQTLTEKAKLEBEKQISDASRQSLRDLDAEAKKQVEKOLANT 226
QY 304 EKHIDKVFHKDLOQLOVDKALQQAQEMLEKABE-RHQBKDFLLKAEVSOBKMLKQ 362
DB 227 AE-LDKVGEDK-----QISDASRQSLRDLDAEAKKQVEKOL------ANLTAE 273
QY 363 QETHLQQLALYTEKEFEFONTLKSSEVFTTFKQEMKMTKKIKLEKETTYRSMES 422
DB 274 -----VKEKQISDASRQSLRDLDAEAKKQVEKALEANSKLALEK----- 318
QY 423 SNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQOTERNDLNKQVQDSAGGQSLTD 482
DB 319 LNK-----ELESEKKTLEKEXAELOAK--LEASAKAL---KEQAKQAEELAKLAKGASD 369
QY 483 S-OPERRP-----EGPQAQAPSSPRVTEAPCPGAPSTEASGTPQEPETSAR 529
DB 370 SQTPDYKPGNKAVPGKQAPQACTKPNQKAPMKETKQPLPSTGETANPFFTAAR 424

RESULT 39

US-09-914-259-21
; Sequence 21, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Loligo pealeii
US-09-914-259-21

Query Match 8.0%; Score 217.5; DB 4; Length 967;
Best Local Similarity 20.0%; Pred. No. 1e-06;
Matches 110; Conservative 95; Mismatches 245; Indels 101; Gaps 17;

QY 48 QSGALRDVSELSROLEDIL--STYCDNNQGGPGEAGGPAEPEDAESKRTTYVARG 105
DB 436 QSOLIEKLEQOMEQ-EDLIQARRDYENLQDMRSIQADNESAKDEVKEVQAL----- 489
QY 106 EPFTPTVYVYGEKPSKDPNTEIROSDEVGDR-----DHR----- 142
DB 490 --BELAMVDQKQVEDNKNENLSBELNQLSTLNLQNELDQLKQSSXHKRVRTD 547
QY 143 -----POEKKKAKGLGKE-----ITLLMQTLTSLTSPK 172

DB 548 MMINLLKDLGDIGTIVGNAAEETKTAGSGEKIBBEFTVARIYISOMKSEVKTLSVRNQ 607
QY 173 LAALCKKTAELLEHRNSQOMKLL-QKKQSQLOVQEKDHLRGEHSAVLAARSKLSLRE 231
DB 608 LENTQDDNFKKIEHEDKLSNCKLLIQHBAKMASLQBAIKDSNKKMLKLENDVDSLNEE 667
QY 232 LQHRNSKLEKGVQARBEKEKKEVTSHTFQVTLNDIOLQMEQHNRNSK----- 281
DB 668 YAK-----LXAEQEMHLAALSEREX-TSQASETREVLEKQEMHREHQKQLQSLRDEIS 722
QY 282 -----LRQENMELAEHLKLLIEQYB-LREHIDKVFHKDLOQLOVDKALQQAQEM- 331
DB 723 EKQATVDMLKDDNQRSLALEKLOADYDKLQEEVEKAAKLADLSLQ-IDRQEQAKQDLK 781.
QY 332 -----LKEAEERQREKDFLLKAEVSOBKMLKQOETHLQQLALYTEKEFEFONTL 385
DB 782 GLEETVALEQLTLEHLKLFVQDLONKVKSCSTSEDEDTGGN-AAQKKQKISPLENNL 840
QY 386 SKSSEVFTTFKQEMKMTKKIKLEKETTYRSMESNKALLEMAEKTVDKLEGLQ 445
DB 841 EQLTKVHQQLVRDNADLRCELPLEKLEKLRATMERVKSDESALKD-AKEGAMRDR--KRYQ 897
QY 446 VKIQRLEKLCRALQOTERNDLNKRVQDLSAGGQSLTDSGPERPEGQAQAPSSPRVTEA 505
DB 898 HEVDRIKEAVKQKMLARGHAAQIAKPIRPGQHSVSPAQAARIRGGGLSQNGPMITST 957
QY 506 PCYCPGAPSTEA 516
DB 958 PIRM-APESKA 967

RESULT 40

US-08-875-435B-4
; Sequence 4, Application US/08875435B
; Patent No. 6593304
; GENERAL INFORMATION:
; APPLICANT: Hasegawa, Kazuhide
; APPLICANT: Arakawa, Emi
; APPLICANT: Oda, Shoji
; APPLICANT: Matsuda, Yuzuru
; APPLICANT: Takahashi, Katsuhito
; APPLICANT: Sugahara, Michihiro
; APPLICANT: Ishiyama, Haruo
; TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
; TITLE OF INVENTION: FOR MYOSIN HEAVY CHAIN SMI ISOFORM PROTEIN INSERTED INTO
; TITLE OF INVENTION: VECTOR DNA, MICROORGANISM CARRYING THE RECOMBINANT DNA, AND
; TITLE OF INVENTION: AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
; FILE REFERENCE: 07898-013001
; CURRENT APPLICATION NUMBER: US/08/875,435B
; CURRENT FILING DATE: 1997-07-25
; PRIOR APPLICATION NUMBER: PCT/JP96/00134
; PRIOR FILING DATE: 1996-01-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-08-875-435B-4

Query Match 8.0%; Score 217.5; DB 4; Length 1972;
Best Local Similarity 20.4%; Pred. No. 2.5e-06;
Matches 111; Conservative 107; Mismatches 184; Indels 143; Gaps 21;
QY 34 QAPRPEGAQARTAGSALRDVSELSROLEDILSTYCDNNQGGPGEAGGPAEPED 93
DB 846 QVTRQEMQAKEDLOKIKERQQAESBELQKQKHTQLSEKNNLLOEQQAETELVAE 905
QY 94 AEKSRITYVARGPEFTPVYVG-----EKEPSKQDP-NTBEIROSDEVGDRHRRPQE-- 145
DB 906 AEMVRVLAAR-KQELEBILHEMARLEEBEDRQQLQAEKQKQAKQMLDLSEQLSEEA 964


```

534 NEEKRQOLQOLHEYETELEDENEPALAAAKKLEGLQKXLELOADSAITKGREAIKQ 593
310 VFK-----HKDLOQLVDAK-----LQAOEMLKEABERHQR- 341
594 LRKLQAKMDFQBELEDAASRDEITPATAKENKKAASLEADLMQLEDLAAABARRKQA 653
342 --EKDFLLKEAVESQRMCLMKQOETHLQKQALYATEKFBEFQNTLSKSSVFTTFKQEM 399
654 DLKXELABELASSLSGRNALQDKERLEARTAOLEEELEBEQGNMEAMSD-----RV 706
400 EKMTKKIKLEKETMYRSWESSNKALLEM-ABEKTVRDK-----ELEG-----L 444
707 RKATQQAQELSNLATERSTAQGNESAPQOLERQRNKLKSLHEMEGAVSKPKFTIAAL 766
445 QVKIQLEKLCRALQTERNDLNRKV 470
767 EAKIAQLEQVEQBEAREKQAATKSLK 792

```

RESULT 45
 US-08-685-871-2
 ? Sequence 2, Application US/08685871
 ? Patent No. 6013499
 ? GENERAL INFORMATION:
 ? APPLICANT: NARUMIYA, Shuh
 ? APPLICANT: IWANATSU, Akihiro
 ? TITLE OF INVENTION: RHO TARGET PROTEIN KINASE P160
 ? NUMBER OF SEQUENCES: 68
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Foley & Lardner
 ? STREET: 3000 K Street, N.W., Suite 500
 ? CITY: Washington
 ? STATE: D.C.
 ? COUNTRY: USA
 ? ZIP: 20007-5109
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patent In Release #1.0, Version #1.30
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/685,871
 ? FILING DATE: 24-JUL-1996
 ? CLASSIFICATION: 435
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: JP 8-184102
 ? FILING DATE: 25-JUN-1996

```

PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 7-262553
FILING DATE: 14-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16887/845
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-635-871-2

Query Match      8.0%; Score 215.5; DB 3; Length 1354;
Best Local Similarity 25.2%; Pred. No. 2.2e-06;
Matches 101; Conservative 64; Mismatches 145; Indels 91; Gaps 16;

127 BEIRQSDVGDHRPQEKKAKGIGKEITILMOTLNTLSTPEEKALCKKCYAELDEE 186

```

487	Db	EKMLLOHRIINERYQKRAQEKNEKRNENEVSTLTKOOLEDL	-----KKVSQ----	531
187	Oy	HRNSQOKMMLQKQKSOQLVQKQKHLRGEHSKAVLAR	-----SKLESTCRBLQRHN	236
532	Db	-----NSQLANEKLSQLOKQLEEANDLLRTESPTAVLRLEKSHTEKSKSIQSLSENLRLBQRN	-----	589
237	Oy	RSLEKEGVQRARE	-----EBEKR-----KGVIGHFOVTLNDILOQME--QHN--E	277
590	Db	RILENSKSKQTDKQYVYQLOQILAEERDRDGRCHDSEMGLOARITSLQESVYKHLKXNLEKVE	-----	649
278	Oy	RNSKLRQENMELARLKLLEQVELREHHDKVFKHDKLOOQL	-----VDKLOQA	328
650	Db	GERKBAQDMLNHSKEKXNNLE	-----IDLNYLKLSQORLEQSVNNEHKVTKARLTDK	701
323	Oy	QEMLKAEERHQREKDFLLKAEVBSQRMCMLMQOETHLKQALALYTERFEFQNTLSKS	-----	388
702	Db	HQSTEEAKSVAMCEMKLKEEREAREKAE	-----NRVVQIEKQCSMLDVLKQSQKQL--	755
339	Oy	SEVETTFRQEMEKTKKIK-KLEKBTTWYSRSHSSNKALLEWABEKT--	-----VDEKLEGLQ	445
756	Db	-----EHLTGNKERNMEDEVKNTLTQLEQ-----	-----SNKRLLOQNEUKTOAFRADNLIKGLE	804
446	Oy	-----VKIQRLSKLQALQTERNDLNKRVDLSAGGQGSUTD	482	
805	Db	QOMKQEIINTLLEAKRLLEFFSLAQTLKOYR-----GNQGMRE	841	

```

RESULT 46
US-09-976-594-296
; Sequence 296, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 296
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 064987CD1
US-09-976-594-296

```

Query Match	8.0%	Score 215;	DB 4;	Length 1388;
Best Local Similarity	21.3%;	Pred. No. 2.4e-06;		
Matches 132;	Conservative 84;	Mismatches 173;	Indels 230;	Gaps 23;
QY	91	PEDAQRS	-----TVVARNGEPE-----	---TPVV---113
DB	316	PEDARISKAANNLICAPLTDREVLRGNVGEIRIQHPFFKNDQHWNIRETAAPVVP	EL 375	
QY	114	-----YCEKPSKGDNPTEI-----	-----RQSDVVG 136	
DB	376	SSDIDSSNFDDIEDDKGVDVETPIPKAFVGNQLPFGFTYYRKNLLSDSPSCRENDSIQ	435	
QY	137	DRDHRAPQE-----	-----KKKAGLQKEITLLMQTL 163	
DB	436	SRKNESQETQKKLYLIEHLSNEMOAKELEBKCKSVNTRLEKATKELEEITLRKSV	E 495	
QY	164	NTLSTPBEKLAALCKKYAEL-----	-----LEEHRNSOK-QMKLQK- 200	
DB	496	SALRQLERERVALLQHNAAEYQKADHEADKRNLENDVNSLKQLEDLKKNQNSQISTE	555	
QY	201	-----OSQLVQEKDHLRGESKAV-LARSKLES--LCRELQRNRSLKEE-	242	

SEQUENCE CHARACTERISTICS:
LENGTH: 1388 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-695-576-4

Query Match 7.9%; Score 214; DB 2; Length 1388;
Best Local Similarity 20.5%; Pred. No. 2.8e-06;
Matches 127; Conservative 87; Mismatches 175; Indels 230; Gaps 21;

QY 91 PEDAEKSR-----TVVARGSEPE-----PTPVV--- 113
DB PEDAEISGHAKMLICAFITDREVLGRNGVTEIRQHPFKNDQDQHNIRETAPVVPPEL 375
QY 114 -----YGRKPSKGPDPNTEI-----RQSDVVG 136
DB SSDSDSNFDDIEDDKGVETFPFKAFVGNQLPFGTYYVRENLLSDSPSCRENDISQ 435
QY 137 DRDHRPQE-----KKKAGLGKEITLLMQTL 163
DB SRKNESEIIOKKLYTLEHLSNEMQAELEBQKCKSVNTRLEKTAKELEEEITLRKSV 495
QY 164 NTLSTPEKLAALCKKYAE-----LLEHRNSQKQMK 195
DB SALQRLEREKALLQHKNAEYQKADHEADKKRNLENDVNSLKQLEDLKKGNQNSQISTE 555
QY 196 LQKQSQSVQEKDHLRGEHSAV-LABS KLES--LCRELQHNRSKLEE----- 242
DB KYNQLOQLDSTNALLRTSTAAARKTQAEBSKIQOLESNNRDLQDNKCLLETAKLG 615
QY 243 -----GVCRAREEEKEKVTSHFQVTLNDIQ-----LQMEQH 275
DB LEKEPFIQSALESERRDR---THGSEINDLQGRICGLEBDLNGKILLAKVEKRQL 672
QY 276 NERNSKLQOE--NVELABRLKLIHOVELREHIDKVFVKHDLQOQLVD-----A 323
DB QERPTDLKEKSNBEIDMTYQLKVIQOQSLEGEAE--HKATKARLADNKNKIYESTEEA 728
QY 324 KLOQAQAEMLKE-ABERHOREK-DFLLKEAVS-----SORMCELMKQOETELKQ 369
DB KSEAKEMEKLLBERTLKQKVENMLLEABKRCSLDCLDKSQSOQKINELLKQDVLNE 787
QY 370 QLALYTERPE-----FQNTLSKSEVFTT-----FQEMKMKTKIKKLEKETT 415
DB DVRLNLTKEIQTOKRCLTQNDLQKQVNTLQKQVNTLQKQVNTLQKQVNTLQKQVNT 847
QY 416 YRSRWESSNKALLEWAE-----KTVRDKELSGLQVKIQLRLEKLCRALQTEINDLNR 468
DB LRKERQDAGQMKELQDLEAEQYFSTYKTQVRELKEECEEKTKGKELQKQKQELQDE 907
QY 469 VQDLSAGQGSGLTDSGPER 487
DB RDSLAQAQLEITLTKADSEQ 926

RESULT 48
US-08-714-741-40
Sequence 40, Application US/08714741
Patent No. 6500613
GENERAL INFORMATION:
APPLICANT: Briles, David E.
APPLICANT: McDaniel, Larry S.
APPLICANT: Swiatlo, Edwin
APPLICANT: Yotter, Janet
APPLICANT: Crain, Marilyn J.
APPLICANT: Hollingshead, Susan
APPLICANT: Tart, Rebecca
APPLICANT: Brooks-Walter, Alexis
TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
TITLE OF INVENTION: PORTIONS AND PRODUCTS
NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:
ADDRESSER: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,741
FILING DATE: 16-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer Esq., William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2460
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 864 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-714-741-40

Query Match 7.9%; Score 213; DB 4; Length 864;
Best Local Similarity 22.4%; Pred. No. 1.8e-06;
Matches 120; Conservative 99; Mismatches 207; Indels 110; Gaps 23;

QY 34 QAPRKEGQAQTAQGAIRDVSEELSR-----QLEDILSTYCVDDNNQGGP---GEQGAQG 86
DB QAEAEVESQAEATRLKIKITDREAKRKADAKLKEAVEKNVATSEQDKPKRRKRGVSG 206
QY 87 BPAREP-----DAKSTYVARGPEPTPVVYGEKPSGDPNTEIROSDVGDHRR 142
DB ELATPKKENDAKSSDSSVGEETLPSPLNMANESQ-----TEHKVDVEY----- 252
QY 143 PQBKKAAGLGEITLLMQTLNTLSTPEKLAALCKKYAELISEHNSQKMKLQKQS 202
DB -----IKKMLSEIQDRRKHTQVNLNTKLGAITKLYLYELSVLKENSKEELTSKTKA 306
QY 203 QLVQ-----EKDHLRGEHSAVLAESKLSLCELOHNRSLKEGVQVARE--BEKQK 255
DB ELTAAEQFQKOTLKEP-KKVAEAKV-----EEAKKAKQKQKEDRN 350
QY 256 EVTSHFQVTLNDIQLOMEQHNRNSKLRQNNMELA-----ERLKKLIQYELREE 305
DB -----TVKTLLEBIA--ESDVKVAEAELELVKEANESENRBEKIQAQKEKVESKKA 403
QY 306 HDKVFHKDLOQOLVDKLOQAQAEMLKEAEERHQR--EKDFLLKEAV-----ESQRMCE 358
DB EATRLKIKITDREKKAEEAKRBAESEKKAQAKQKVDABEYALAEKIAELSEYEVORLEK 463
QY 359 LAKQ-----QETHLKOOL--ALYTEKPEEFQNTLSKSESVFTTFQKQEMKMTKIKKLEKE 412
DB ELKEIDSESDSELYLKEGLEAPLQSKLDTKKAKLSLEEL-----SDKIDELDAEIAKLEVQ 519
QY 413 TTMYRSRWESSNKALLEWAEKTVRDKELEGLQVKIQLRLEKLCRALQTEINDLNGVQDL 472
DB --LQDAEGNNVVEAFYFKEGLEKTTAPKKAEE-----LEKAEADLKKAVDPR 562
QY 473 SAGGQGSGLTDSGPERRPEGPAQSPSPRVTPAPCPVGPAPSEASQCTGQPTSA 528
DB ETAPAPQAPAPAEKPAEKP-APAEKPA-APAEKPA-APAEKPA-PAPEKAPA 613

```
RESULT 49
US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 6645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; TITLE OF INVENTION: Chromosome Congression
; FILE REFERENCE: 18557C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/059,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2954
; TYPE: PRT
; ORGANISM: Xenopus sp.
; FEATURES:
; OTHER INFORMATION: Xenopus centromere-associated protein-X (XCENP-X)
; OTHER INFORMATION: member of the kinesin superfamily of microtubule
; OTHER INFORMATION: motor proteins
; FEATURES:
; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
; US-09-150-867-1

Query Match 7.8%; Score 212; DB 4; Length 2954;
Best Local Similarity 22.2%; Pred. No. 9.9e-06;
Matches 127; Conservative 83; Mismatches 169; Indels 192; Gaps 25;

QY 52 LRDVSEL--SRQLEILSTYCVANNQ--GPGDGAQGPAPEDAEKSRITYVARNGEP 107
DB 1284 LRAAQBELREQQLVDSFRQQLDCSVGISPNHD----- 1318

QY 108 EPTPVVYGEKPSKGDPTNTEIRQSDV--GDRDHRPQEK----- 146
DB 1319 ----AVANQEKVSLGEVNS---LQSEMLRGERDELQTSCKALVSELELLRAHVKSVEGEN 1371

QY 147 ----KKAAGLGEITLLMOTLNTLSTPBEKLAALCKYABILLSEHRNSQKMKLIQ--KK 200
DB 1372 LEITKGLNGLEKEILLGKSESEVILKSVLENKEDNNKLEQAEEYSKKNQPSLEEVFSG 1431

QY 201 QSOLOVEKDLRGEHSKAVIARSLKSLCRS---LQRENRSI-----KEEG 243
DB 1432 SQKLVDEIEVLKQ--LKA--AEERLEIKDRIFELVGTANTVNEVGEKLETPLOADHEEDS 1488

QY 244 VQAREEERK--KEYTSHFQVTLNDIQLOMGEHNRNSKLR--QENME-----LAB 291
DB 1489 IDRRSEEMBIKVLGKELRNQVLLERLQ--BEKLELSNKLBITQKEMETSVLLKDDLQ 1545

QY 292 RLKLLIEQVELREHDKVFK--HKDLQQQLVDKALQQAQEMLKE----- 334
DB 1546 KLESLSSENIILKENTDTTLKHSDTQAQL--QKTQDELQALANLAIASDNCPTQKEK 1603

QY 335 -----AEERHQK-----EKDFLKEAVESQR 355
DB 1604 TSADCVHPELEKILLUTBELHQTNEQKLLHKEBLEQAQVELKCEVEHLKMSIESKS 1663

356 MCELMKQOETH-----LQQALALYTEKEPEEFONT-----LSKSEVFTTFKQ 397
DB 1664 SLESIL-QHEKHDTEQQLLALQQQVVTQBEKKELQQTHEHLTAEDVHLKENIELGLNFKN 1722

QY 398 EMERKTKK-----IKKLEKETTMYRSRWESSKAL-----LEMAEKTVR-DKE 440
DB 1723 EAQQTTKBEQCLLNENKELQSQHRLQCEIBELMKSLADKESALEFTLKSQBQKVINLQOE 1782

QY 441 LEGQVKIORLEKLCRALQTERNDLNKRQVD 471
DB 1783 MEMWMEBELKNSORTVIAERDQLQDOLRE 1813

RESULT 50
US-09-914-259-27
; Sequence 27, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1027
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-914-259-27

Query Match 7.8%; Score 211.5; DB 4; Length 1027;
Best Local Similarity 23.4%; Pred. No. 2.9e-06;
Matches 117; Conservative 87; Mismatches 223; Indels 73; Gaps 17;

QY 47 AQSGLRDVSELRSQLEILSTYCVNNQGPBGDGAQGPAPEDAEKSRITYVARNGE 106
DB 475 SENDAKDEVKEVLQALELAVNY--DQKSEVTEKSSQONQLLVDELQKVTATMLSELE 532

QY 107 PEPTPVVYGEKPSKGDPTNTEIRQSDS---VGDRDHRPQEKKAAGLGEITLLMOT 162
DB 533 LQRLQEVSGHQKRIAEVLNGLMRDLSEFVSIVGNGBIKLPVEISGA--IBEEFTVARLY 590

QY 163 LNTLSTPPEKLAALCKYAEI--LEEHRNSQKMKLIQKQSQOLVQEKDHLRG--RHSKAV 219
DB 591 ISKIKSEVKSVMKRCRQLENLQVECHRRQGEVVTGRELSSCQLLISQHEAKISLITEYMTV 650

QY 220 -LARSKLE---SLCRSELQ-----HNRSLKEEGVQARAEERKEKKEVTSHFQVTLND 267
DB 651 ELKRRHLEESYDLSDELARLQAHTVHEVALKDKEPDTQDAEEVKA----- 698

QY 268 IQLOMGEHNRNSKLRQENMELARLKKLIEQVELREHIDKVPKHDLOQ--QLVDAKL 325
DB 699 LELQENHREAHHR-----QLARLDRINEKQKTIDEL---KDLNQKLQLELEKL 745

QY 326 QQAQEMLKEAB-----ERHOREKDFL--LKEAV--ESORMCBLMKQOETHL 367
DB 746 QADYERLKNVEENKSAKQELTFLYERHEQSDQLKGLSETVARELQTLHNLRLKLFVQDV 805

QY 368 KQOALALYTEKEPEEFONTLSKSESVFTTFKQEMERKTKKIKKLEKETMYRSRWESSKAL 427
DB 806 TTRVKSAAEMEPESGGTHSQKQKISFLNNEQLQTKVHKQLVRDNADRLCPKLEKRL 865

QY 428 LEMAEKTVRKDEEG-LQVKIORLEKLCRALQTERNDLNKRQVDSLQAGGSLTDSGPE 486
DB 866 RATAE-----RVKALEGALKKEAGKMDKRRYQQEVDRIKEAVRYKSSKGGHSAQIAKP 921

QY 487 RRPFGQAQAPSSPRVTEAP 506
DB 922 VRPGHYPASSPTNPGTRSP 941
```

Search completed: June 7, 2004, 14:42:05
Job time : 40 secs

107 1.3 1064 1 CARB_BACST
108 1.3 1065 1 CARB_BACCL
109 1.3 1071 1 DAM1_MOUSE
110 1.3 1078 1 DAM1_HUMAN
111 1.3 1079 1 IR2P_SCHPO
112 1.3 1085 1 YAPA_SCHPO
113 1.3 1139 1 SR22_CRIGR
114 1.3 1141 1 SR22_HUMAN
115 1.3 1159 1 SR22_HUMAN
116 1.3 1191 1 TOP2_ASFB2
117 1.3 1192 1 TOP2_ASFB7
118 1.3 1372 1 RPOC_RICCN
119 1.3 1372 1 RPOC_RICPR
120 1.3 1453 1 Y373_BOVIN
121 1.3 1498 1 G0A3_HUMAN
122 1.3 1657 1 I0G1_HUMAN
123 1.3 1657 1 I0G1_MOUSE
124 1.3 1940 1 MYH3_HUMAN
125 1.3 1940 1 MYH3_RAT
126 1.3 2144 1 BP28_HUMAN
127 1.3 3034 1 CLRL1_MOUSE
128 1.3 4568 1 DHHB_CHLRE
129 1.3 5171 1 BPEA_HUMAN
130 1.3 7389 1 BPA1_MOUSE
131 1.3 8797 1 SNEL1_HUMAN
132 1.3 49 1 LHB2_RHOCA
133 1.3 56 1 YDAE_ECOLI
134 1.3 66 1 SCXL_ANDAU
135 1.3 68 1 V07K_FMV
136 1.3 71 1 EXTS_STRAS
137 1.3 75 1 R17E_THEVO
138 1.3 76 1 RPOZ_GUITH
139 1.3 77 1 RL29_MYCBO
140 1.3 77 1 RL29_MYCTU
141 1.3 79 1 EXTS_ECOLI
142 1.3 79 1 EXTS_SALTY
143 1.3 79 1 TATA_CAMJE
144 1.3 80 1 AXA7_MOUSE
145 1.3 83 1 RS27_HUMAN
146 1.3 83 1 RS27_XENLA
147 1.3 83 1 VTB3_AGR75
148 1.3 83 1 YN78_RHIME
149 1.3 89 1 CH10_PORGI
150 1.3 91 1 RRI19_DEA
151 1.3 91 1 YAIN_ECOLI
152 1.3 92 1 Y247_CHLCV
153 1.3 93 1 GATC_HELPJ
154 1.3 93 1 GATC_HELPY
155 1.3 95 1 HNGY_MOUSE
156 1.3 96 1 Y143_BORBU
157 1.3 96 1 Y157_METUA
158 1.3 99 1 GATC_FALSO
159 1.3 99 1 Y088_UREPA
160 1.3 100 1 Y667_CHLMU
161 1.3 103 1 SPGB_HUMAN
162 1.3 103 1 SPGB_PANTR
163 1.3 105 1 TMM4_VACGS
164 1.3 105 1 Y082_YEAST
165 1.3 106 1 HNGI_CRIGR
166 1.3 106 1 HNGI_HUMAN
167 1.3 106 1 YTH5_RHOER
168 1.3 109 1 CDK1_MOUSE
169 1.3 109 1 YNEM_PVMR
170 1.3 109 1 YNEM_PVMR
171 1.3 110 1 YHBJ_ACTAC
172 1.3 112 1 Y0DB_BACSU
173 1.3 113 1 PTWX_ECOLI
174 1.3 114 1 CDK1_MESAU
175 1.3 115 1 CALR_PHYBI
176 1.3 115 1 CDK1_HUMAN
177 1.3 117 1 GUPK_HALME
178 1.3 118 1 YNIF_AZOER
179 1.3 121 1 Y879_HALN1

180 6 1.1 123 1 PFDE_AERPE
181 6 1.1 126 1 RS11_METAC
182 6 1.1 126 1 RS11_METMA
183 6 1.1 126 1 SYN2_MOUSE
184 6 1.1 127 1 MINE_CHLUV
185 6 1.1 127 1 NB7M_BOVIN
186 6 1.1 127 1 RS11_HALN1
187 6 1.1 128 1 RR9_ASTIO
188 6 1.1 128 1 RS11_HALMA
189 6 1.1 129 1 RS11_METJA
190 6 1.1 130 1 RS11_METTH
191 6 1.1 130 1 RS11_THEAC
192 6 1.1 130 1 VG67_BPH2
193 6 1.1 131 1 CAIF_ECOLI
194 6 1.1 131 1 GC5H_XANCP
195 6 1.1 131 1 GC5H_XANCP
196 6 1.1 131 1 RS11_AERPE
197 6 1.1 132 1 CYCP_PARSF
198 6 1.1 132 1 PFDA_PYRAE
199 6 1.1 132 1 RS11_THEVO
200 6 1.1 133 1 NL21_PARJU
201 6 1.1 133 1 NL22_PARJU
202 6 1.1 133 1 RS11_ARCFU
203 6 1.1 134 1 ACP1_BRANA
204 6 1.1 134 1 ACP2_BRANA
205 6 1.1 134 1 ACP3_BRANA
206 6 1.1 134 1 ACP_BACAM
207 6 1.1 134 1 ATPF_FUSNN
208 6 1.1 135 1 Y1G4_METJA
209 6 1.1 136 1 ACP2_ARATH
210 6 1.1 136 1 ACP3_ARATH
211 6 1.1 136 1 Y914_METJA
212 6 1.1 137 1 RS11_METKA
213 6 1.1 137 1 RS11_PYRFU
214 6 1.1 137 1 RS11_PYRHO
215 6 1.1 138 1 NUSE_PHOLL
216 6 1.1 138 1 NUSE_YERPE
217 6 1.1 139 1 RS14_SCHPO
218 6 1.1 141 1 HBA2_TACAC
219 6 1.1 141 1 RS12_PLAFA
220 6 1.1 141 1 TCL6_HUMAN
221 6 1.1 141 1 XKDS_BACSU
222 6 1.1 141 1 ZRAP_ECOLI
223 6 1.1 144 1 RS14_TYRBB
224 6 1.1 144 1 SODM_ARCFL
225 6 1.1 144 1 Y047_ARCFU
226 6 1.1 145 1 TH13_CORNE
227 6 1.1 145 1 Y1A1_ECOLI
228 6 1.1 146 1 CDD_MOUSE
229 6 1.1 147 1 YRAK_BACSU
230 6 1.1 148 1 DUT_RICCN
231 6 1.1 148 1 DUT_RICPR
232 6 1.1 148 1 RL9_THETH
233 6 1.1 150 1 PMCD_PSEAE
234 6 1.1 150 1 PDAD_METKA
235 6 1.1 150 1 YVBA_BACSU
236 6 1.1 152 1 ADOM_BPT3
237 6 1.1 152 1 MGSA_SALTY
238 6 1.1 153 1 HOXT_AZOVI
239 6 1.1 153 1 YB34_COXBU
240 6 1.1 154 1 Y712_RICPR
241 6 1.1 155 1 RR7_GUSEU
242 6 1.1 155 1 YHBI_HCMVA
243 6 1.1 157 1 Y53C_AGR75
244 6 1.1 157 1 Y935_AQUAE
245 6 1.1 158 1 Y612_RHIME
246 6 1.1 159 1 FMF5_ECOLI
247 6 1.1 163 1 DTD_NEMTB
248 6 1.1 165 1 MSRA_UREPA
249 6 1.1 165 1 RS16_CAUCR
250 6 1.1 165 1 RUVA_RHIME
251 6 1.1 166 1 NHAX_BACSU
252 6 1.1 166 1 VDEL_BPP4

O9yc11 aeropyrum p
O8trr0 methanosarc
O8pv17 methanosarc
O64332 mus musculus
P56350 chlorella v
O22387 bos taurus
Q9hgj5 halobacteri
P58135 astasia lon
P10788 haloarcula
P54021 methanococc
O26143 methanobact
O9hjd8 thermoplasm
P16517 bacterioph
Q47081 escherichia
P6138 xanthomonas
Q8p6c9 xanthomonas
Q9yb55 aeropyrum p
P00143 paracoccus
O8zt69 pyrobaculum
Q97b94 thermoplasm
P55958 parietaria
O04401 parietaria
O28001 archaeoglob
P10352 brassica na
P17650 brassica na
P32897 brassica na
P07088 brassica ca
Q8rgs3 fusobacteri
Q80303 methanococc
P25702 arabidopsis
Q58324 methanococc
O8tbv9 methanopyru
O8u063 pyrococcus
Q59304 pyrococcus
O7n0j0 photorhabdu
Q6zc42 yeersinia pe
O14150 schizosacch
P01978 tachygllossu
O97249 plasmodium
P56846 homo sapien
P54338 bacillus su
P32682 escherichia
P19800 trypanosoma
P28766 parastichop
O30189 archaeoglob
P52228 corynebacte
P11287 escherichia
P56389 mus musculu
P50862 bacillus su
Q92174 rickettsia
P27151 thetramus the
P17837 pseudomonas
O8txd4 methanopyru
P37503 bacillus su
P07693 bacterioph
Q8xer5 salmonella
P30781 azotobacter
Q83br9 coxiella bu
Q9zcl5 rickettsia
P46292 cuscuta eur
P09702 human cytom
Q8uhx2 agrobacteri
O67076 aquifex ae
Q92s95 rhizobium m
P13719 escherichia
Q9k143 neisseria m
Q9pck2 ureaplasma
P58122 caulobacter
Q92ql0 rhizobium m
O07552 bacillus su
P12551 bacterioph

253	6	1.1	167	1	FIXE_RHIME	P18397 rhizobium m	RT	*Identification of a cDNA for a human high-molecular-weight B-cell growth factor."
254	6	1.1	167	1	Y491_METJA	Q57914 methanococc	RT	Proc. Natl. Acad. Sci. U.S.A. 90:6330-6334(1993).
255	6	1.1	168	1	Y224_HABIN	P44580 haemophilus	RL	[2]
256	6	1.1	168	1	YD02_METJA	Q58698 methanococc	RN	
257	6	1.1	169	1	DSB2_PSEAE	P57701 pseudomonas	RP	ERRATUM
258	6	1.1	170	1	R1MM_XANCP	Q8b0c2 xanthomonas	RX	MEDLINE=96353961; PubMed=8755619;
259	6	1.1	170	1	R1MM_XYLF	Q9ph38 xylella fas	RA	Ambrus J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,
260	6	1.1	170	1	R1MM_XYLF	Q87f55 xylella fas	RA	Tamayo A., Claypool K., McCourt D., Srikiatchatochorn A., Ford R.J.,
261	6	1.1	171	1	YF87_METJA	Q58982 methanococc	RA	Proc. Natl. Acad. Sci. U.S.A. 93:8154-8154(1996).
262	6	1.1	171	1	ADXH_DROME	P37193 drosophila	RN	[3]
263	6	1.1	172	1	IF3_CAUCR	Q9a9d9 caulobacter	RP	SEQUENCE OF 130-259 FROM N.A.
264	6	1.1	173	1	DEF_RHIME	Q92sh6 rhizobium m	RC	TISSUE=Placenta;
265	6	1.1	174	1	RPC1_BPMU	P06019 bacterioph	RX	MEDLINE=22388257; PubMed=12477932;
266	6	1.1	175	1	AXIK_ARATH	Q24410 arabidopsis	RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
267	6	1.1	175	1	BADR_RHOPA	O07458 rhodospheud	RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
268	6	1.1	175	1	SSB_SERMA	P25762 serratia ma	RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
269	6	1.1	177	1	NUOE_RICPR	Q9zdh5 rickettsia	RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
270	6	1.1	178	1	N12M_HUMAN	Q9y6m9 homo sapien	RA	Diatchenko L., Marushina K., Farmer A.A., Rubin G.M., Hong L.,
271	6	1.1	178	1	PRYR_ENTFA	O52707 enterococc	RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Sheetz T.E.,
272	6	1.1	179	1	GRPE_LACIA	Q9cgv9 lactococcus	RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
273	6	1.1	179	1	GRPE_LACIC	P42369 lactococcus	RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
274	6	1.1	179	1	GRPE_STRMU	O06941 streptococc	RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
275	6	1.1	179	1	R1MM_MYCLE	Q33016 mycobacteri	RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
276	6	1.1	179	1	YMD8_SALTY	Q9zq29 salmonella	RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
277	6	1.1	180	1	KDOP_HABIN	P45314 haemophilus	RA	Fahy J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
278	6	1.1	180	1	Y576_AQUAE	O67453 aquifex aeo	RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
279	6	1.1	181	1	HE28_HUMAN	Q13442 homo sapien	RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
280	6	1.1	182	1	ARL3_HUMAN	P36405 homo sapien	RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
281	6	1.1	183	1	ASPH_AERHI	P31735 aeromonas h	RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
282	6	1.1	184	1	AROK_CHLMU	P47593 mycoplasma	RT	Schmerch A., Schein J.B., Jones S.J.M., Marra M.A.,
283	6	1.1	184	1	IPYR_MYCGE	P75250 mycoplasma	RT	"Generation and initial analysis of more than 15,000 full-length
284	6	1.1	184	1	IPYR_MYCFN	P75250 mycoplasma	RT	human and mouse cDNA sequences."
285	6	1.1	185	1	RFP_THETN	O8rae2 thermocoe	CC	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
286	6	1.1	185	1	RFP_HELHP	Q7v1i0 helicobacte	CC	-1- CAUTION: The clone described in Ref.1 was initially thought to
287	6	1.1	185	1	YHQO_YEAST	P38812 saccharomyc	CC	code for a high molecular weight interleukin (IL-14). The
288	6	1.1	186	1	ATPD_CYAPA	P48082 cyanophora	CC	translation of this cDNA does not predict the open reading frame
289	6	1.1	186	1	ATPF_MESVI	Q9mut1 mesostigma	CC	for the 60 kDa protein described in Ref.1. A reading frame on the
290	6	1.1	186	1	COAE_THERH	O56416 thermus the	CC	plus strand predicts a 7.7 kDa protein. The longest open reading
291	6	1.1	186	1	RRF_BABEA	O8rt64 bartonella	CC	frame (shown here) is on the opposite strand and predicts a 36.4
292	6	1.1	186	1	YRKQ_BACSU	P54430 bacillus su	CC	kDa protein. The relationship of this sequence to IL-14, if any,
293	6	1.1	187	1	Y4GA_RHLSN	P55457 rhizobium s	CC	is uncertain.
294	6	1.1	189	1	R13A_SALTR	Q91487 salmo trutt	CC	-----
295	6	1.1	189	1	RK12_SPIOL	P02398 spinacia ol	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
296	6	1.1	190	1	GRPE_STRPY	Q99yc8 streptococc	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
297	6	1.1	192	1	CDRC_CLOAB	Q45812 clostridium	CC	the European Bioinformatics Institute. There are no restrictions on its
298	6	1.1	192	1	INAI_RAT	P05011 rattus norv	CC	use by non-profit institutions as long as its content is in no way
299	6	1.1	194	1	DMOI_DESMO	P21505 desulfuroco	CC	modified and this statement is not removed. Usage by and for commercial
300	6	1.1	195	1	PYRE_SULSO	Q9ux09 sulfolobus	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
							CC	or send an email to license@isb-sib.ch).
							CC	-----
							DR	EMBL; LI5344; -; NOT ANNOTATED CDS.
							DR	EMBL; BC029686; AAH29686.1; -;
							DR	PIR; A48203; A48203.
							DR	Genew; HGNC:5976; IL14.
							DR	MTM; 147684; -;
							DR	GO; GO:0005576; C:extracellular; NAS.
							DR	GO; GO:0030372; F:high molecular weight B-cell growth factor . . .; NAS.
							DR	GO; GO:0008283; P:cell proliferation; NAS.
							KW	Hypothetical protein.
							SQ	SEQUENCE 259 AA; 30107 MW; 8AB8D1AA6891357C CRC64;
							QY	Query Match 48.9%; Score 259; DB 1; Length 259;
							Best Local Similarity 100.0%; Pred. No. 3.6e-250; Indels 0; Gaps 0;	
							Matches 259; Conservative 0; Mismatches 0;	
							QY	272 MEQHNRNSKLRQENNELAERLKKLIEQYELAREEHIDKVFKEHDKQQQLVDAKLOQAQRM 331
							DB	1 MEQHNRNSKLRQENNELAERLKKLIEQYELAREEHIDKVFKEHDKQQQLVDAKLOQAQRM 60
							QY	332 LKEAERHQREKDFLLKEAVSRQRCMLKQOETHLKOOLALYTERKFEFQNTLSKSSV 391
							DB	61 LKEAERHQREKDFLLKEAVSRQRCMLKQOETHLKOOLALYTERKFEFQNTLSKSSV 120

ALIGNMENTS

RESULT 1	YL14_HUMAN	STANDARD;	PRT;	259 AA.
YL14_HUMAN	Q8NRY3;			
AC	P40222; Q8NRY3;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Hypothetical protein initially thought to be identical with			
DE	interleukin-14 (IL-14) (High molecular weight B-cell growth factor)			
DE	(HMM-BCGF).			
GN	IL14.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=9337675; PubMed=8327514;			
RA	Ambrus J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,			
RA	Tamayo A., Claypool K., McCourt D., Srikiatchatochorn A., Ford R.J.,			

QY 392 FTTFQEMKMTKKIKLEKFTMYRSWESSNKALLEMAEKTVRDKLEGLQVKIQL 451
Db 121 FTTFQEMKMTKKIKLEKFTMYRSWESSNKALLEMAEKTVRDKLEGLQVKIQL 180
QY 452 EKLCALQTERDLNKRVDLSAGGSGISITDSGPERPEGCAQAPSSPRVTEAPCPYGA 511
Db 181 EKLCALQTERDLNKRVDLSAGGSGISITDSGPERPEGCAQAPSSPRVTEAPCPYGA 240
QY 512 PSTEASGOTGPOEPTSA 530
Db 241 PSTEASGOTGPOEPTSA 259
RESULT 2
ID FLO8_YEAST STANDARD; PRT; 799 AA.
AC P40067; Q05751;
DF 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcriptional activator FLO8 (PDS5 protein).
GN FLO8 OR PDS5 OR YER105C/YER108C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Sukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sigma 12789;
RA Liu H., Styles C.A., Fink G.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 71-799 FROM N.A.
RC STRAIN=ATCC 60715;
RX MEDLINE=96335146; PubMed=8757402;
RA Kobayashi O., Suda H., Ohtani T., Some H.;
RT "Molecular cloning and analysis of the dominant flocculation gene
FLO8 from Saccharomyces cerevisiae.";
RL Mol. Gen. Genet. 251:707-715(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313264; PubMed=9169868;
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
Hunkle-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
Oh C., Petel P.X., Roberts D., Sehl P., Schramm S., Shogren T.,
Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V";
RL Nature 387:78-81(1997).
CC -!- FUNCTION: Required for diploid filamentous growth, haploid
invasive growth and flocculation. Putative transcriptional
activator of FLO1.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to a
frameshift.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).
CC
CC EMBL; U51431; AAC49522.1; -;
DR EMBL; D83713; BRA12076.1; -;
DR EMBL; U18916; AAC03207.1; ALT_FRAME.
DR EMBL; U18916; AAC03206.1; ALT_FRAME.
DR GenOnline; 139198; -;

DR TRANSPAC; T03311; -;
DR SGD; S0000911; FLOS.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR InterPro; IPR006594; Lish.
DR SMART; SM00667; Lish; 1.
DR PROSITE; PS00896; Lish; 1.
KW Transcription regulation; Activator; Nuclear protein.
FT DOMAIN 73 105 LISH.
FT DOMAIN 41 55 POLY-GLN.
FT CONFLICT 112 112 V -> I (IN REF. 2).
FT CONFLICT 115 115 P -> S (IN REF. 2).
FT CONFLICT 383 383 C -> G (IN REF. 2).
FT CONFLICT 441 441 A -> T (IN REF. 2).
FT CONFLICT 447 447 A -> V (IN REF. 2).
FT CONFLICT 598 598 R -> P (IN REF. 2).
SQ SEQUENCE 799 AA; 86834 MW; 91AAE10D8E586DAF CRC64;
Query Match 1.7%; Score 9; DB 1; Length 799;
Best Local Similarity 100.0%; Pred. No. 1.8; Mismatches 0; Gaps 0;
Matches 9; Conservative 0; Indels 0;
Qy 164 NTLSTPEEK 172
Db 671 NTLSTPEEK 679
RESULT 3
CARB_METAC STANDARD; PRT; 1070 AA.
ID CARB_METAC
AC Q8TNY4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
phosphate synthetase ammonia chain).
OS CARB OR MA2143.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
FitzRugh W.E., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
Allen N., Naylor J., Staeger-Thomann N., DeArelano K., Johnson R.,
Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,
Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
and physiological diversity.";
RL Genome Res. 12:532-542(2002).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
promotes the hydrolysis of glutamine to ammonia, which is used by
the large (or ammonia) chain to synthesize carbamoyl phosphate (By
similarity).
CC -!- SIMILARITY: Belongs to the carb family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 479-500.
RC STRAIN=New Zealand white; TISSUE=Brain;
RX MEDLINE=93374895; PubMed=8396129;
RA Lamphear B.J., Yan R., Yang P., Waters D., Liebig H.-D.,
RT Klump H., Kuechler E., Skern T., Rhoads R.E.;
RT "Mapping the cleavage site in protein synthesis initiation factor
RT eIF-4 gamma of the 2A proteases from human Coxsackievirus and
RT rhinovirus.";
RL J. Biol. Chem. 268:19200-19203(1993).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=93054654; PubMed=1429670;
RA Yan R., Rychlik W., Etchison D., Rhoads R.E.;
RT "Amino acid sequence of the human protein synthesis initiation factor
RT eIF-4 gamma.";
RL J. Biol. Chem. 267:23226-23231(1992).
CC -!- FUNCTION: Component of the protein complex EIF4F, which is
CC involved in the recognition of the mRNA cap, ATP-dependent
CC unwinding of 5'-terminal secondary structure and recruitment of
CC mRNA to the ribosome.
CC -!- SUBUNIT: EIF4F is a multi-subunit complex, the composition of
CC which varies with external and internal environmental conditions.
CC It is composed of at least EIF4A, EIF4E and EIF4G. EIF4G interacts
CC with the serine/threonine kinases MNK1 and MNK2. Appears to act
CC as a scaffold protein, holding these enzymes in place to
CC phosphorylate EIF4E (By similarity).
CC -!- SIMILARITY: THE C-TERMINAL REGION IS SIMILAR TO THE N-TERMINAL
CC REGION OF WHEAT EURARYOTIC INITIATION FACTOR (ISO)4P SUBUNIT P82.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L22090; AAA31242.1; -;
DR PIR; I46707; I46707.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR003307; eIF5C.
DR InterPro; IPR003890; IF_eIF4G.
DR InterPro; IPR003891; IF_eIF4G_MA3.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF02847; MA3; 1.
DR Pfam; PF02854; MIF4G; 1.
DR Pfam; PF02020; W2; 1.
DR SMART; SM00515; eIF5C; 1.
DR SMART; SM00544; MA3; 1.
DR SMART; SM00543; MIF4G; 1.
KW Initiation factor; Protein biosynthesis; Translation regulation;
FT Phosphorylation; RNA-binding.
FT DOMAIN 188 192 POLY-PRO.
FT DOMAIN 262 275 POLY-GLU.
FT DOMAIN 1393 1398 POLY-GLU.
SQ SEQUENCE 1402 AA; 154050 MW; 7FD85D7E30519230 CRC64;

Query Match 1.7%; Score 9; DB 1; Length 1402;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 SEELSRQLE 64
Db 1243 SEELSRQLE 1251

RESULT 6
ID SYE DEIRA STANDARD; PRT; 483 AA.
AC Q9RZ30;
DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)
DE (Glurs).
GN GLTX OR DR0485.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Moffat R.S., Hatt D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Dodson K.J., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC diphosphate + L-glutamyl-tRNA(Glu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A5001907; AAF10063.1; ALT_INIT.
DR PIR; H75514; H75514.
DR HSSP; P27000; 1GLN.
DR TIGR; DR0485; -;
DR HAMAP; MF_00022; -; 1.
DR InterPro; IPR004527; Gltx_bact.
DR InterPro; IPR000924; Glu_tRNA-synt_1c.
DR InterPro; IPR008925; tRNA-synt_bind.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR PRINTS; PR00987; TRNASYNTHGLU.
DR TIGRPFAMs; TIGR00464; gltx_bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 14 24 "HIGH" REGION.
FT SITE 253 257 "RMSKS" REGION.
FT BINDING 256 256 ATP (BY SIMILARITY).
SQ SEQUENCE 483 AA; 54689 MW; 7FG6F5C0BF3A4567 CRC64;

Query Match 1.5%; Score 9; DB 1; Length 483;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 496 AFSSPRVT 503
Db 3 AFSSPRVT 10

RESULT 7
ID MURD RICCN STANDARD; PRT; 500 AA.
AC Q92I60;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)
 UNP-N-acetylmuramoyl-L-alanine-D-glutamate ligase (EC 6.3.2.9) (UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase) (D-glutamic acid adding enzyme).
 MURD OR RC0560.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RA MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RA "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 RL Science 293:2093-2098(2001).
 CC -!- FUNCTION: Cell wall formation. Catalyzes the addition of glutamate to the nucleotide precursor UDP-N-acetylmuramoyl-L-alanine (UMA) (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanine + Glutamate = ADP + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamate.
 CC -!- PATHWAY: Peptidoglycan biosynthesis.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the murCDEF family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: AE008617; ALN03098.1; -
 DR PIR: H97269; H97769.
 DR HAMAP: MF_00639; atypical; 1.
 DR InterPro: IPR000713; Mur_ligase.
 DR InterPro: IPR004101; Mur_ligase_C.
 DR InterPro: IPR005762; MurD.
 DR Pfam: PF01225; Mur_ligase; 1.
 DR Pfam: PF02875; Mur_ligase_C; 1.
 DR TIGRFAMs: TIGR01087; murD; 1.
 DR Peptidoglycan synthesis; Cell wall; Cell division; Ligase;
 KW ATP-binding; Complete proteome.
 FT NP_BIND 111 117 ATP (POTENTIAL).
 FT DOMAIN 260 306 RPE3.
 SQ SEQUENCE 500 AA; 55787 MW; 024B268F6074822C CRC64;
 Query Match 1.5%; Score 8; DB 1; Length 500;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 85 QGSPAEPE 92
 Db 278 QGSPAEPE 285
 ID -CALD RAT STANDARD; PRT; 531 AA.
 AC Q62736;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Non-muscle caldesmon (CDW) (L-caldesmon).
 GN CALD1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;

[1]
 RN SEQUENCE FROM N.A., MUTAGENESIS, AND PHOSPHORYLATION SITES.
 RC TISSUE=Liver;
 RX MEDLINE=95181370; PubMed=7876150;
 RA Yamashiro S., Yamakita Y., Yoshida K.-S., Takiguchi K., Matsumura F.;
 RT "Characterization of the COOH terminus of non-muscle caldesmon mutants lacking mitosis-specific phosphorylation sites";
 RL J. Biol. Chem. 270:4023-4030(1995).
 RN [2]
 RN PHOSPHORYLATION BY CDC2.
 RC MEDLINE=91095023; PubMed=1986309;
 RA Yamashiro S., Yamakita Y., Hosoya H., Matsumura F.;
 RT "Phosphorylation of non-muscle caldesmon by p34cdc2 kinase during mitosis";
 RL Nature 349:169-172(1991).
 CC -!- FUNCTION: Actin- and myosin-binding protein implicated in the regulation of actomyosin interactions in smooth muscle and nonmuscle cells (could act as a bridge between myosin and actin filaments). Stimulates actin binding of tropomyosin which increases the stabilization of actin filament structure. In muscle tissues, inhibits the actomyosin ATPase by binding to P-actin. This inhibition is attenuated by calcium-calmodulin and is potentiated by tropomyosin. Interacts with actin, myosin, two molecules of tropomyosin and with calmodulin. Also play an essential role during cellular mitosis and receptor capping.
 CC -!- SUBCELLULAR LOCATION: On thin filaments in smooth muscle and on stress fibers in fibroblasts (nonmuscle) (By similarity).
 CC -!- TISSUE SPECIFICITY: High-molecular-weight caldesmon (h-caldesmon) is predominantly expressed in smooth muscles, whereas low-molecular-weight caldesmon (l-caldesmon) is widely distributed in non-muscle tissues and cells. Not expressed in skeletal muscle or heart (By similarity).
 CC -!- DOMAIN: The N-terminal part seems to be a myosin/calmodulin-binding domain, and the C-terminal a tropomyosin/actin/calmodulin-binding domain. These two domains are separated by a central helical region in the smooth-muscle form.
 CC -!- PTM: In non-muscle cells, phosphorylation by CDC2 during mitosis causes caldesmon to dissociate from microfilaments.
 CC Phosphorylation reduces caldesmon binding to actin, myosin, and calmodulin as well as its inhibition of actomyosin ATPase activity. Phosphorylation also occurs in both quiescent and dividing smooth muscle cells with similar effects on the interaction with actin and calmodulin and on microfilaments reorganization.
 CC -!- SIMILARITY: Belongs to the caldesmon family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: U18419; AAA68521.1; -
 DR PIR: A55887; A55887.
 DR InterPro: IPR006017; Caldesmon.
 DR InterPro: IPR006018; Caldesmon_LSP.
 DR Pfam: PF02029; Caldesmon; 1.
 DR PRINTS: PR01076; CALDESMON.
 DR Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;
 KW Alternative splicing.
 FT DOMAIN 20 200 MYOSIN AND CALMODULIN-BINDING (BY SIMILARITY).
 FT DOMAIN 303 360 TROPOMYOSIN-BINDING (POTENTIAL).
 FT DOMAIN 402 412 TROPOMYOSIN-BINDING (POTENTIAL).
 FT DOMAIN 392 424 STRONG ACTIN-BINDING (BY SIMILARITY).
 FT DOMAIN 454 460 CALMODULIN-BINDING (BY SIMILARITY).
 FT DOMAIN 506 531 WEAK ACTIN-BINDING (BY SIMILARITY).
 FT DOMAIN 33 40 POLY-ARG.
 FT DOMAIN 180 189 POLY-GLU.
 FT DOMAIN 279 282 POLY-ARG.
 FT DOMAIN 319 322 POLY-GLU.

FT DOMAIN 336 339 POLY-GLU.
 FT MOD_RES 249 249 PHOSPHORYLATION (BY CDC2).
 FT MOD_RES 462 462 PHOSPHORYLATION (BY CDC2).
 FT MOD_RES 468 468 PHOSPHORYLATION (BY CDC2).
 FT MOD_RES 491 491 PHOSPHORYLATION (BY CDC2).
 FT MOD_RES 497 497 PHOSPHORYLATION (BY CDC2).
 FT MOD_RES 527 527 PHOSPHORYLATION (BY CDC2).
 FT MUTAGEN 249 249 S->A: DECREASES STRONGLY PHOSPHORYLATION-DEPENDENT ACTIN BINDING.
 FT MUTAGEN 462 462 S->A: DECREASES PHOSPHORYLATION-DEPENDENT ACTIN BINDING.
 FT MUTAGEN 468 468 T->A: DECREASES PHOSPHORYLATION-DEPENDENT ACTIN BINDING.
 FT MUTAGEN 491 491 S->A: DECREASES PHOSPHORYLATION-DEPENDENT ACTIN BINDING.
 FT MUTAGEN 497 497 S->A: DECREASES PHOSPHORYLATION-DEPENDENT ACTIN BINDING.
 FT MUTAGEN 527 527 S->A: DOES NOT DECREASE PHOSPHORYLATION-DEPENDENT ACTIN BINDING.
 FT SEQUENCE 531 AA; 60584 MW; CBCEC50271A23829 CRC64;
 Query Match 1.5%; Score 8; DB 1; Length 531;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 247 AREEEER 254
 DB 334 AREEEER 341
 RESULT 9
 ID_VG50 HSVSA STANDARD; PRT; 535 AA.
 AC 001012;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE Probable transcription activator EDRF1.
 GN 50 OR HDR21.
 OS Herpesvirus saimiri (strain 11).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=10383;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92333688; PubMed=1321287;
 RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
 RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
 RA Honess R.N.;
 RT "Primary structure of the herpesvirus saimiri genome.";
 RL J. Virol. 66:5047-5058(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230228; PubMed=1314457;
 RA Nicholas J., Coles L.S., Newman C., Honess R.W.;
 RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
 RT herpesvirus saimiri (HVS) L-DNA: General conservation of genetic
 RT organization between HVS and Epstein-Barr virus.";
 RL Virology 188:296-310(1992).
 RN [3]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RA Nicholas J., Coles L.S., Newman C., Honess R.W.;
 RT "Regulation of the herpesvirus saimiri (HVS) delayed-early
 RT 110-kilodalton promoter by HVS immediate-early gene products and a
 RT homolog of the Epstein-Barr virus R trans activator.";
 RL J. Virol. 62:2457-2466(1988).
 CC -!- FUNCTION: TRANSCRIPTION ACTIVATION. REGULATES THE DELAYED-EARLY
 CC 110 KDA PROMOTER.
 CC -!- SIMILARITY: TO EBV BRFL1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EXBL; X64346; CAA45672.1; ALT INIT.
 DR EXBL; M86409; AAA46124.1; ALT INIT.
 DR EXBL; M60850; AAA46159.1; ALT INIT.
 DR InterPro; IPR004998; Herpes_TA50.
 DR Pfam; PF03326; Herpes_TA50_1.
 KW Transcription regulation; Activator; DNA-binding; Early protein.
 SQ SEQUENCE 535 AA; 60050 MW; B4F2B9ABA38816FB CRC64;
 Query Match 1.5%; Score 8; DB 1; Length 535;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 151 GLGKEITL 158
 DB 187 GLGKEITL 194
 RESULT 10
 ID_TFC5 YEAST STANDARD; PRT; 594 AA.
 AC P46678;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transcription factor TFIIB B" component (TFIIB90).
 GN TFC5 OR TFC7 OR YNL039W OR N2682.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=9288C / AB972;
 RX MEDLINE=96003864; PubMed=7568218;
 RA Kassavetis G.A., Nguyen S.T., Kobayashi R., Kumar A.,
 RA Geiduschek E.P., Pisanò M.;
 RT "Cloning, expression, and function of TFC5, the gene encoding the B'
 RT component of the Saccharomyces cerevisiae RNA polymerase III
 RT transcription factor TFIIB.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:9786-9790(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9288C;
 RX MEDLINE=96203116; PubMed=8617241;
 RA Rueth J., Conesa C., Dieci G., Lefebvre O., Dueterhoeft A.,
 RA Ottolenghi S., Sentenac A.;
 RT "A suppressor of mutations in the class III transcription system
 RT encodes a component of yeast TFIIB.";
 RL EMBO J. 15:1941-1949(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9288C;
 RA Roberts S., Miller S., Lane W.S., Hahn S.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Dueterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D., Hilbert H.,
 RA Moestl D.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: General activator of RNA polymerase III transcription.
 CC -!- SUBUNIT: TFIIB comprises the TATA-binding protein (TBP), the
 CC B-related factor (BRF) and the B' component (TFC5).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: TO S.POMBE SPCC1919.14C.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; U31819; AAC49073.1; -;
 EMBL; U38415; AAC49364.1; -;
 EMBL; U37533; AAC49348.1; -;
 EMBL; Z71315; CA95906.1; -;
 PIR; S62141; S62141. -;
 GERMOnline; 143046; -;
 SGD; S0004984; TFC5.
 InterPro: IP001005; Myb DNA-binding.
 Pfam: PF00249; myb_DNA-binding; 1.
 SMART; SM00717; SANT; 1.
 KW Transcription regulation; Activator; Nuclear protein.
 SQ SEQUENCE 594 AA; 67687 MW; 7B0115BBB2451175 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 594;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 249 EEEKRKE 256
 D 334 EEEKRKE 341

RESULT 11
 YP58_YEAST STANDARD; PRT; 756 AA.
 ID Q99293;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein YPL158C.
 GN YPL158C OR F2570.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=9710377; PubMed=8948103;
 RA Purnelle B., Coster F., Goffeau A.;
 RT "The sequence of 55 kb on the left arm of yeast chromosome XVI
 identifies a small nuclear RNA, a new putative protein kinase and two
 new putative regulators.";
 RL Yeast 12:1483-1492(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=97313271; PubMed=9169875;
 RA Bussey H., Storms R.K., Ahmed A., Albertmann K., Allen E., Ansoorge W.,
 RA Araujo R., Aparicio A., Barrall B.G., Badcock K., Benes V.,
 RA Brustin D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
 RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
 RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoft A.,
 RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
 RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
 RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Klein K.,
 RA Komp C., Kurdi O., Laekari D., Lew H., Lin A., Lin D., Louis E.J.,
 RA Marathe R., Messinghoff F., Meses H.-W., Mitiapati S., Moestl D.,
 RA Mueller-Auer S., Namath A., Newnich U., Oefner P., Pearson D.,
 RA Patel F.X., Pohl T.M., Purnelle B., Schafer M., Scharfe M.,
 RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
 RA Urestrazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
 RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
 RA Zhong W.N., Zollner A., Vo D.H., Hani J.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI";
 RL Nature 387:103-105(1997).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; X96770; CAA85563.1; -;
 EMBL; Z73514; CAA97863.1; -;
 PIR; S65169; S65169.
 GERMOnline; 144140; -;
 SGD; S0006079; YPL158C.
 KW Hypothetical protein; Coiled coil.
 FT DOMAIN 253 263
 FT DOMAIN 648 711 COILED COIL (POTENTIAL).
 FT DOMAIN 652 658 POLY-GLU.
 FT DOMAIN 664 669 POLY-GLU.
 FT DOMAIN 686 690 POLY-GLU.
 SQ SEQUENCE 758 AA; 84845 MW; 3B4FA92891C87F2B CRC64;

Query Match 1.5%; Score 8; DB 1; Length 758;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 249 EEEKRKE 256
 D 687 EEEKRKE 694

RESULT 12
 RPOB_LOTJA STANDARD; PRT; 1070 AA.
 ID Q9BB59;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6).
 GN RPOB.
 OS Lotus japonicus.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.
 OX NCBI_TaxID=34305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Accession MG-20;
 RX MEDLINE=21082923; PubMed=11214967;
 RA Kato T., Kaneko T., Sato S., Nakamura Y., Tabata S.;
 RT "Complete structure of the chloroplast genome of a legume, Lotus japonicus.";
 RL DNA Res. 7:323-330(2000).
 CC -1- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).
 CC -1- SUBUNIT: In chloroplasts, the RNA polymerase is composed of four subunits: alpha, beta, beta', and beta".
 CC -1- SUBCELLULAR LOCATION: Chloroplast.
 CC -1- SIMILARITY: Belongs to the RNA polymerase beta chain family.
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; AP002983; BAB3194.1; -;
 HSSP; Q9KNU7; 1HQW.

```

DR InterPro; IPR007121; RNA_pol_B.
DR InterPro; IPR007644; RNA_pol_Rpb2_1.
DR InterPro; IPR007642; RNA_pol_Rpb2_2.
DR InterPro; IPR007645; RNA_pol_Rpb2_3.
DR InterPro; IPR007120; RNA_pol_Rpb2_6.
DR InterPro; IPR007641; RNA_pol_Rpb2_7.
DR Pfam; PF04563; RNA_pol_Rpb2_1; 1.
DR Pfam; PF04561; RNA_pol_Rpb2_2; 1.
DR Pfam; PF04565; RNA_pol_Rpb2_3; 1.
DR Pfam; PF00562; RNA_pol_Rpb2_6; 1.
DR Pfam; PF04560; RNA_pol_Rpb2_7; 1.
DR PROSITE; PS01166; RNA_POL_BETA; 1.
KW Transferrase; DNA-directed RNA polymerase; Transcription; Chloroplast.
SQ SEQUENCE 1070 AA; 120892 MW; 7126BD99B0B5C063 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 1070;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 ESLCRELQ 233
DB 241 ESLCRELQ 248

RESULT 13
CARB_THETN
ID CARB_THETN STANDARD; PRT; 1072 AA.
AC Q8REK0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR TWE0816.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
SEQUENCE FROM N.A.
STRAIN=MB4 / JCM 11007;
MEDLINE=219292816; PubMed=1197336;
BAO Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
Tan H., Chen R., Wang J., Yu J., Yang H.;
RA "A complete sequence of T. tengcongensis genome.";
RT Genome Res. 12:689-700(2002).
RL CC
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COPACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -!- SIMILARITY: Belongs to the carb family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE013048; AM24073.1; -.
CC HAVAP; MF_01210; -.
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR005483; CPhase_L.
CC InterPro; IPR005479; CPhase_L_D2.
CC InterPro; IPR005480; CPhase_L_D3.
CC InterPro; IPR005481; CPhase_L_N.

```

```

DR InterPro; IPR004362; MGS-like.
DR Pfam; PF00289; CPhase_L_Chain; 2.
DR Pfam; PF02786; CPhase_L_D2; 2.
DR Pfam; PF02787; CPhase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPhase.
DR TIGRfam; TIGR01369; CPhaseII_lrg; 1.
DR PROSITE; PS00866; CPhase_1; 1.
DR PROSITE; PS00867; CPhase_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 545 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 547 929 ALLOSTERIC DOMAIN.
FT REPEAT 930 1072
FT REPEAT 547 1072
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1072 AA; 119000 MW; D5F08ADC6BE75D37 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 1072;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 SLKEGVQ 245
DB 36 SLKEGVQ 43

RESULT 14
TRAA_AGRTS
ID TRAA_AGRTS STANDARD; PRT; 1100 AA.
AC Q44349;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Conjugal transfer protein traa.
GN TRAA OR ATU6127 OR AGR_PTI_237.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
CG Plasmid pTiC58.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=96312367; PubMed=8763953;
RA Farrand S.K., Hwang I., Cook D.M.;
RA "The tra region of the nopaline-type Ti plasmid is a chimera with
RT elements related to the transfer systems of RSP1010, RP4, and F.";
RL J. Bacteriol. 178:4233-4247(1996).
RN [2]
SEQUENCE FROM N.A.
MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.P. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayin T., Levy R., Li M.-J., McClelland E., Palmieri A., Gordon D.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Perry M.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [3]

```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Ouello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlet K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA William C., Allinger M., Doughty D., Scott C., Lappas C., Markez B.,
RA Planagan C., Crowell C., Gursen J., Lomo J., Sear C., Strub G.,
RA Cielo C., Slater S.,
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
CC -!- SIMILARITY: Belongs to the mobA/mobL family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AF010180; AAC17212.1; -.
DR EMBL; AF009431; AAL46363.1; -.
DR EMBL; AE007939; AAX91091.1; ALT_INIT.
DR PIR; AE3243; AE3243.
DR PIR; T03419; T03419.
DR InterPro; IPR005053; MobA_MobL.
DR Pfam; PF03389; MobA_MobL; 1.
KW Conjugation; ATP-binding; Plasmid; Complete proteome.
FT NP_BIND 404 411
FT CONFLICT 372 372 L -> LL (IN REF. 1).
SQ SEQUENCE 1100 AA; 123477 MW; B85D31DB526B7344 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 1100;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 135 VGDHRHR 142
DB 671 VGDHRHR 678
|||||

RESULT 15
MY10 BOVIN
ID MY10 BOVIN STANDARD; PRT; 2052 AA.
AC P79114.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myosin X.
GN MYO10.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Artera;
RC MEDLINE=20442246; PubMed=10984435;
RA Berg J.S., Derfler B.H., Pennisi C.M., Corey D.P., Cheney R.E.;
RT "Myosin-X, a novel myosin with pleckstrin homology domains, associates
RT with regions of dynamic actin."
RL J. Cell Sci. 113:3439-3451(2000).
CC -!- FUNCTION: Myosins are actin-based motor molecules with ATPase
CC activity. Unconventional myosins serve in intracellular movements.
CC Their highly divergent tails are presumed to bind to membranous
CC compartments, which would be moved relative to actin filaments (By
CC similarity). Plays a role in regions of dynamic actin.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 FERM domain.
CC -!- SIMILARITY: Contains 3 IQ domains.
CC -!- SIMILARITY: Contains 1 MYTH4 domain.

CC -!- SIMILARITY: Contains 2 PH domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; U55042; AAB39486.1; -.
DR PIR; T18519; T18519.
DR HSPF; P08799; LMND.
DR InterPro; IPR000299; Band 4.1.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR000857; MYTH4.
DR InterPro; IPR001849; PH.
DR Pfam; PF00612; IQ; 3.
DR Pfam; PF00663; myosin_head; 1.
DR Pfam; PF00784; MYTH4; 1.
DR Pfam; PF00169; PH; 3.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00295; B41; 1.
DR SMART; SM00015; IQ; 3.
DR SMART; SM00242; MISC; 1.
DR SMART; SM00139; MYTH4; 1.
DR SMART; SM00233; PH; 2.
DR PROSITE; PS00660; FERM_1; FALSE NEG.
DR PROSITE; PS00661; FERM_2; FALSE NEG.
DR PROSITE; PS00057; FERM_3; 1.
DR PROSITE; PS50096; IQ; 2.
DR PROSITE; PS50003; PH_DOMAIN; 2.
KW Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat.
FT DOMAIN 1 727 MYOSIN HEAD-LIKE.
FT DOMAIN 742 763 IQ 1.
FT DOMAIN 764 787 IQ 2.
FT DOMAIN 788 817 IQ 3.
FT DOMAIN 1206 1304 PH 1.
FT DOMAIN 1386 1491 PH 2.
FT DOMAIN 1694 2038 FERM.
FT DOMAIN 800 941 COILED COIL (POTENTIAL).
SQ SEQUENCE 2052 AA; 235837 MW; 43DF13424B4B2D28 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 2052;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 248 REEEKRKX 255
DB 824 REEEKRKX 831
|||||

RESULT 16
RK21 ODOSI
ID RK21 ODOSI STANDARD; PRT; 105 AA.
AC P49557;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Chloroplast 50S ribosomal protein L21.
GN RPL21.
OS Odontella sinensis (Marine centric diatom).
OC Chloroplast.
OC Eukaryota; Stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC Biddulphiophycidae; Eupodiscaceae; Odontella.
OX NCBI_TaxID=2839;
RN [1]
RP SEQUENCE FROM N.A.
RA Kowalik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;
RT "The chloroplast genome of a chlorophyll a+c-containing alga,
RT Odontella sinensis.";
```

```
RL Plant Mol. Biol. Rep. 13:336-342(1995).
CC -!- FUNCTION: This protein binds to 23S ribosomal RNA in the presence
CC of protein L20 (by similarity).
CC -!- SIMILARITY: Belongs to the L21P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z67753; CAA91663.1; -
CC PIR; S78290; S78290.
CC InterPro; IPR001787; Ribosomal_L21P.
CC Pfam; PF00829; Ribosomal_L21P; 1.
CC ProDom; PD003604; Ribosomal_L21P; 1.
CC TIGRfam; TIGR00061; L21; 1.
CC PROSITE; PS01169; RIBOSOMAL_L21; 1.
CC Ribosomal protein; rRNA-binding; Chloroplast.
CC SEQUENCE 105 AA; 12428 MW; ED99D758B93ECD008 CRC64;
CC -----
CC Query Match 1.3%; Score 7; DB 1; Length 105;
CC Best Local Similarity 100.0%; Pred. No. 30;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 152 LGKEITL 158
CC |||||
CC DB 30 LGKEITL 36
CC -----
RESULT 17
INSC_ECOLI
ID INSC ECOLI STANDARD; PRT; 121 AA.
AC P19776; O07989; O08018; O08019; P76357; P77346;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transposase insc for insertion element IS2A/D/F/H/I/K.
GN (INSC1 OR B0360) AND (INSC2 OR B1403) AND (INSC3 OR B1997) AND
GN (INSC4 OR B2861) AND (INSC5 OR B3044) AND (INSC6 OR B4272).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=88137965; PubMed=2830172;
RA Ronscher H.J., Rak B.;
RT "Genetic organization of insertion element IS2 based on a revised
RT nucleotide sequence.";
RL Gene 59:291-296(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT The complete genome sequence of Escherichia coli K-12.;
RT Science 277:1453-1474(1997).
```

```
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M., Mori T.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampel G., Seki Y., Sivasundaram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampel G., Seki Y.,
RA Sivasundaram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -!- FUNCTION: Involved in the transposition of the insertion sequence
CC IS2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V00279; CAA23542.1; -
CC EMBL; U14003; AAA97168.1; ALT INIT.
CC EMBL; U28377; AAA69212.1; ALT INIT.
CC EMBL; U28375; AAA83043.1; ALT INIT.
CC EMBL; AE000143; AAC73463.1; ALT INIT.
CC EMBL; AE000237; AAC74485.1; ALT INIT.
CC EMBL; AE000291; AAC75058.1; ALT INIT.
CC EMBL; AE000369; AAC75900.1; -
CC EMBL; AE000386; AAC76080.1; -
CC EMBL; AE000498; AAC77228.1; ALT INIT.
CC EMBL; D90778; BAA15013.1; ALT INIT.
CC EMBL; D90779; BAA15019.1; ALT INIT.
CC EMBL; D90838; BAA15822.1; ALT INIT.
CC EMBL; D90850; BAA16005.1; ALT INIT.
CC EMBL; D90851; BAA16013.1; ALT INIT.
CC EMBL; D90852; BAA16036.1; ALT INIT.
CC EcoGene; EG40003; insc.
CC InterPro; IPR002514; Transposase_8.
CC Pfam; PF01527; Transposase_8; 1.
CC Complete proteome.
CC VARIANT 34 34 L -> F (IN B1997).
CC SEQUENCE 121 AA; 13452 MW; 59431E5C452E067A CRC64;
CC -----
CC Query Match 1.3%; Score 7; DB 1; Length 121;
CC Best Local Similarity 100.0%; Pred. No. 34;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY * 346 LLKEAVE 352
CC |||||
CC DB 96 LLKEAVE 102
CC -----
RESULT 18
INSC_SHIFL
```

ID INSC SHIFL STANDARD; PRT; 121 AA.
 AC PS944;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE Transposase INSC for insertion element IS2.
 GN (INSC1 OR SF0245) AND (INSC2 OR SF0879) AND (INSC3 OR SF0933) AND
 GN (INSC4 OR SF0960) AND (INSC5 OR SF1054) AND (INSC6 OR SF1165) AND
 GN (INSC7 OR SF1343) AND (INSC8 OR SF1463) AND (INSC9 OR SF1587) AND
 GN (INSC10 OR SF2011) AND (INSC11 OR SF2615) AND
 GN (INSC12 OR SF2694) AND (INSC13 OR SF2873) AND
 GN (INSC14 OR SF2984) AND (INSC15 OR SF3431) AND
 GN (INSC16 OR SF3512) AND (INSC17 OR SF3805) AND
 GN (INSC18 OR SF3873) AND (INSC19 OR SF3988) AND
 GN (INSC20 OR SF4097) AND (INSC21 OR SF4185).
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Shigella.
 OX NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=301 / Serotype 2a;
 RX MEDLINE=22272406; PubMed=12384590;
 RA Jin O., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 RT through comparison with genomes of Escherichia coli K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441(2002).
 CC -1- FUNCTION: Involved in the transposition of the insertion
 CC sequence IS2 (by similarity).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC ENBL; AE015060; AAM41906.1; ALT INIT.
 DR ENBL; AE015116; AAM42511.1; ALT INIT.
 DR ENBL; AE015123; AAM42562.1; ALT INIT.
 DR ENBL; AE015125; AAM42588.1; ALT INIT.
 DR ENBL; AE015133; AAM42676.1; ALT INIT.
 DR ENBL; AE015143; AAM42781.1; ALT INIT.
 DR ENBL; AE015160; AAM42947.1; ALT INIT.
 DR ENBL; AE015170; AAM43060.1; ALT INIT.
 DR ENBL; AE015180; AAM43173.1; ALT INIT.
 DR ENBL; AE015218; AAM43556.1; ALT INIT.
 DR ENBL; AE015277; AAM44112.1; ALT INIT.
 DR ENBL; AE015284; AAM44187.1; ALT INIT.
 DR ENBL; AE015301; AAM44359.1; ALT INIT.
 DR ENBL; AE015311; AAM44465.1; ALT INIT.
 DR ENBL; AE015351; AAM44892.1; ALT INIT.
 DR ENBL; AE015359; AAM44970.1; ALT INIT.
 DR ENBL; AE015387; AAM45245.1; ALT INIT.
 DR ENBL; AE015394; AAM45310.1; ALT INIT.
 DR ENBL; AE015405; AAM45422.1; ALT INIT.
 DR ENBL; AE015418; AAM45521.1; ALT INIT.
 DR ENBL; AE015427; AAM45606.1; ALT INIT.
 DR InterPro; IPR002514; Transposase 8.
 DR Pfam; PF01527; Transposase 8; 1_
 KW Transposable element; Transposition; DNA-binding; DNA recombination;
 KW Complete proteome.
 SQ SEQUENCE 121 AA; 13452 MW; 59431ES0452E067A CRC64;
 Query Match 1.3%; Score 7; DB 1; Length 121;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 346 LLKEAVE 352
 DB 95 LLKEAVE 102
 |||||
 |||||
 RESULT 19
 CYTC RAT
 ID -CYTC RAT STANDARD; PRT; 127 AA.
 AC P14841;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cystatin C precursor (Fragment).
 GN CST3.
 GN Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Buffalo;
 RX MEDLINE=90092122; PubMed=2689174;
 RA Cole T., Dickson P.W., Esnard F., Averill P., Risbridger G.,
 RA Gauthier F., Schreiber G.;
 RT "The cDNA structure and expression analysis of the genes for the
 RT cysteine proteinase inhibitor cystatin C and for beta 2-microglobulin
 RT in rat brain";
 RL Eur. J. Biochem. 186:35-42(1989).
 RN [2]
 RP SEQUENCE OF 8-127.
 RX MEDLINE=90380276; PubMed=2400577;
 RA Esnard F., Esnard A., Faucher D., Capony J.-P., Derancourt J.,
 RA Brillard M., Gauthier F.;
 RT "Rat cystatin C: the complete amino acid sequence reveals a site for
 RT N-glycosylation";
 RL Biol. Chem. Hoppe-Seyler 371:161-166(1990).
 RN [3]
 RP SEQUENCE OF 8-49.
 RX MEDLINE=89313020; PubMed=3044831;
 RA Esnard A., Esnard F., Faucher D., Gauthier F.;
 RT "Two rat homologues of human cystatin C";
 RL FEBS Lett. 236:475-478(1988).
 RN [4]
 RP SEQUENCE OF 8-20.
 RC TISSUE=Sertoli cells;
 RX MEDLINE=9225121; PubMed=1563513;
 RA Esnard A., Esnard F., Guillou F., Gauthier F.;
 RT "Production of the cysteine proteinase inhibitor cystatin C by rat
 RT Sertoli cells";
 RL FEBS Lett. 300:131-135(1992).
 CC -1- FUNCTION: As an inhibitor of cysteine proteinases, this protein is
 CC thought to serve an important physiological role as a local
 CC regulator of this enzyme activity. Known to inhibit cathepsin B,
 CC H, and L.
 CC -1- SIMILARITY: Belongs to the cystatin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X16957; CAA34831.1; -.
 DR EMBL; S07085; S07085.
 DR EMBL; S10587; S10587.
 DR HSSP; F01034; IG96.
 DR InterPro; IPR000010; Cystatin.
 DR Pfam; PF00031; cystatin; 1.
 DR SMART; SM00043; Cy; 1.
 DR PROSITE; PS00287; CYSTATIN; 1.
 KW Thiol protease inhibitor; Signal.

```
FT NON_TER 1 1
FT SIGNAL <1 7
FT CHAIN 8 127
FT ACT_SITE 18 18
FT SITE 62 66
FT DISULFID 80 90
FT DISULFID 104 124
FT CONFLICT 25 25
SQ SEQUENCE 127 AA; 14039 MW; 78F70158B7525853 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 EGVQRA 247
DB 27 EGVQRA 33

RESULT 20
NL13 PARJU
ID NL13 PARJU STANDARD; PRT; 138 AA.
AC Q40305;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable nonspecific lipid-transfer protein 1 precursor (LTP) (Major
DE pollen allergen Par j 1.0201) (Par j 1) (P1 protein).
OS Parietaria judaica.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Rosales; Urticaceae; Parietaria.
OX NCBI_TaxID=33127;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pollen.
RX MEDLINE=97258596; PubMed=9104790;
RA Duro G., Colombo P., Costa M.A., Izzo V., Porcasi R., di Fiore R.,
RA Locorotondo G., Cocchiara R., Geraci D.;
RT "Isolation and characterization of two cDNA clones coding for
RL Int. Arch. Allergy Immunol. 112:348-355(1997).
CC -!- FUNCTION: Plant nonspecific lipid-transfer proteins transfer
CC phospholipids as well as galactolipids across membranes. May play
CC a role in wax or cutin deposition in the cell walls of expanding
CC epidermal cells and certain secretory tissues.
CC -!- ALLERGEN: Causes an allergic reaction in human. Can bind IgE and
CC induce histamine release from basophils of Pj-pollen-allergic
CC subjects.
CC -!- SIMILARITY: Belongs to the plant LTP family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
FT DISULFID 50 65 BY SIMILARITY.
FT DISULFID 66 111 BY SIMILARITY.
FT DISULFID 86 127 BY SIMILARITY.
SQ SEQUENCE 138 AA; 14414 MW; A4E66230FEC0B6EE CRC64;

Query Match 1.3%; Score 7; DB 1; Length 138;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 EKPSKG 122
DB 58 EKPSKG 64

RESULT 21
NL11 PARJU
ID NL11 PARJU STANDARD; PRT; 139 AA.
AC P43217;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable nonspecific lipid-transfer protein (LTP) (Major pollen
DE allergen Par j 1.0101) (Par j 1) (P5 protein) (Fragment).
OS Parietaria judaica.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Rosales; Urticaceae; Parietaria.
OX NCBI_TaxID=33127;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Flower;
RX MEDLINE=94185791; PubMed=8137937;
RA Costa M.A., Colombo P., Izzo V., Kennedy H., Venturella S.,
RA Cocchiara R., Mistrello G., Palagiani P., Geraci D.;
RT "cDNA cloning, expression and primary structure of Par j1, a major
RT allergen of Parietaria judaica pollen.";
RL FEBS Lett. 341:182-186(1994).
RN [2]
RP REVISIONS TO C-TERMINUS.
RA Geraci D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Plant nonspecific lipid-transfer proteins transfer
CC phospholipids as well as galactolipids across membranes. May play
CC a role in wax or cutin deposition in the cell walls of expanding
CC epidermal cells and certain secretory tissues.
CC -!- ALLERGEN: Causes an allergic reaction in human. Can bind IgE and
CC induce histamine release from basophils of Pj-pollen-allergic
CC subjects.
CC -!- SIMILARITY: Belongs to the plant LTP family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; X77414; CAA54587.1; ALT_INIT.
DR PIR; S77948; S77948.
DR HSP; P19656; LMZM.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR000528; Plant_LTP.
DR Pfam; PF00234; tryp_alpha_amy1; 1.
DR PRINTS; PR00382; LIPIDTRANSF.
DR SMART; SM00499; AAI; 1.
DR PROSITE; PS00597; PLANT_LTP; PARTIAL.
KW Lipid-binding; Transport; Multigene family; Allergen.
FT NON_TER 1 1
FT DISULFID 14 29 BY SIMILARITY.
FT DISULFID 30 75 BY SIMILARITY.
FT DISULFID 50 91 BY SIMILARITY.
```



```

RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
CC -!- FUNCTION: Involved in cell division; probably involved in
CC intracellular septation (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the ispZ family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP005079; BAC60233.1; -.
DR HAMAP; MF 00189; -.
DR InterPro; IPR006008; SeptationA.
DR Pfam; PF04279; IspA; 1.
KW Cell division; Septation; Transmembrane; Complete proteome.
FT TRANSMEM 22 44 Potential.
FT TRANSMEM 51 68 Potential.
FT TRANSMEM 78 100 Potential.
FT TRANSMEM 113 135 Potential.
FT TRANSMEM 150 169 Potential.
SQ SEQUENCE 187 AA; 21317 MW; 23AF8FF42DCBD055 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 187;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LGKEITL 158
DB 104 LGKEITL 110
|||||

RESULT 25
ISPZ_VIBVU STANDARD; PRT; 187 AA.
AC PS9366;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable intracellular septation protein.
GS ISPZ OR VJ3073.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Chey H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Involved in cell division; probably involved in
CC intracellular septation (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the ispZ family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; AE016807; AAO11397.1; ALT_INIT.
DR HAMAP; MF 00189; -.
DR InterPro; IPR006008; SeptationA.
DR Pfam; PF04279; IspA; 1.
DR ProDom; PD016710; SeptationA; 1.
DR TIGRPFAM; TIGR00997; ispZ; 1.
KW Cell division; Septation; Transmembrane; Complete proteome.
FT TRANSMEM 22 44 Potential.
FT TRANSMEM 51 68 Potential.
FT TRANSMEM 78 100 Potential.
FT TRANSMEM 113 135 Potential.
FT TRANSMEM 150 169 Potential.
SQ SEQUENCE 187 AA; 21014 MW; AB803D8B2398A9FC CRC64;

Query Match 1.3%; Score 7; DB 1; Length 187;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LGKEITL 158
DB 104 LGKEITL 110
|||||

RESULT 26
LOLB_NEIMA STANDARD; PRT; 193 AA.
ID _LOLB_NEIMA
AC P57023;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Outer-membrane lipoprotein lolB precursor.
GN LOLB OR NMA1091.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491."
RL Nature 404:502-506(2000).
CC -!- FUNCTION: Plays a critical role in the incorporation of
CC lipoproteins in the outer membrane after they are released by the
CC lola protein (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (By similarity).
CC -!- SIMILARITY: Belongs to the lolB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL162755; CAB84354.1; -.
DR PIR; H81874; H81874.
DR HAMAP; MF 00233; -.
DR InterPro; IPR004565; LolB.
DR ProDom; IPR000437; Prok_lipoprot_S.
DR Pfam; PF03550; LolB; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Chaperone; Outer membrane; Lipoprotein; Transport; Protein transport;
KW Signal; Complete proteome; Palmitate.

```

```
FT SIGNAL 1 15 BY SIMILARITY.
FT CHAIN 16 193 OUTER-MEMBRANE LIPOPROTEIN LOLB.
FT LIPID 16 16 S-diacylglycerol cysteine (By
FT LIPID 16 16 similarity).
FT LIPID 16 16 N-palmitoyl cysteine (By similarity).
SQ SEQUENCE 193 AA; 21156 MW; 7259427B44B4F478 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 BELSROL 63
DB 102 BELSROL 108

RESULT 27
LOLB NEIME STANDARD; PRT; 193 AA.
AC P57024;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Outer-membrane lipoprotein LOLB precursor.
GN LOLB OR NMB0873.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Karchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizze M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
CC -!- FUNCTION: Plays a critical role in the incorporation of
CC lipoproteins in the outer membrane after they are released by the
CC lola protein (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (By similarity).
CC -!- SIMILARITY: Belongs to the lolB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF002439; AAP41284.1; -.
CC TIGR; NMB0873; -.
CC HAMAP; MF_00233; -.
CC InterPro; IPR004565; LOLB.
CC InterPro; IPR000437; Prok_lipoprot_S.
CC Pfam; PF03550; LOLB; 1.
CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Chapterone; Outer membrane; Lipoprotein; Transport; Protein transport;
KW Signal; Complete proteome; Palmitate.
FT SIGNAL 1 15 BY SIMILARITY.
FT CHAIN 16 193 OUTER-MEMBRANE LIPOPROTEIN LOLB.
FT LIPID 16 16 S-diacylglycerol cysteine (By
FT LIPID 16 16 similarity).
```

```
FT LIPID 16 16 N-palmitoyl cysteine (By similarity).
SQ SEQUENCE 193 AA; 21141 MW; 4C2E97734BB2C41E CRC64;
Query Match 1.3%; Score 7; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 BELSROL 63
DB 102 BELSROL 108

RESULT 28
AANT_HDV3 STANDARD; PRT; 195 AA.
ID AANT_HDV3
AC P29996;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Delta antigen.
OS Hepatitis delta virus (isolate D380) (HDV).
OC Viruses; Deltavirus.
OX NCBI_TaxID=31762;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90324949; PubMed=2374010;
RA Saldanha J.A., Thomas H.C., Monjardino J.P.;
RT "Cloning and sequencing of RNA of hepatitis delta virus isolated from
RT human serum."
RL J. Gen. Virol. 71:1603-1606(1990).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTIGEN FROM PATIENTS WITH CHRONIC
CC HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D01075; BAA00874.1; -.
CC PIR; A35219; SAVLH1.
CC InterPro; IPR002506; HDV ag.
CC Pfam; PF01517; HDV ag; 1.
CC ProDom; PD002887; HDV_ag; 1.
KW Antigen; RNA-binding; Nuclear protein; Phosphorylation.
SQ SEQUENCE 195 AA; 21937 MW; 8F09A9C62E9D2721 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45

RESULT 29
AANT_HDV1 STANDARD; PRT; 195 AA.
ID AANT_HDV1
AC P29833;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Delta antigen.
OS Hepatitis delta virus (isolate Lebanon-1) (HDV).
OC Viruses; Deltavirus.
OX NCBI_TaxID=31763;
RN [1]
```

RP SEQUENCE FROM N.A.
RX MEDLINE=92230225; PubMed=1566577;
RA Lee C.M., Bih P.Y., Chao Y.C., Govindarajan S., Lai M.M.C.;
RT "Evolution of hepatitis delta virus RNA during chronic infection.";
RL Virology 188:265-273(1992).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
CC HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M84917; -; NOT_ANNOTATED_CDS.
CC FIR; A40247; SAVILL1.
CC InterPro; IPR002506; HDV_ag.
CC Pfam; PF01517; HDV_ag; 1.
CC ProDom; PD002887; HDV_ag; 1.
KW Antigen; RNA-binding; Nuclear protein; Phosphorylation.
SQ SEQUENCE 195 AA, 21846 MW, 289293C595B5D41B CRC64;

Query Match 1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45

RESULT 30
AA NT HDVM1 STANDARD; PRT; 195 AA.
AC P25881;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Delta antigen.
OS Hepatitis delta virus (isolate Japanese M-1) (HDV).
OC Viruses; Deltavirus.
OX NCBI_TaxID=10424;
RN [1]
RP MEDLINE=91012805; PubMed=2214027;
RA Imazeki F., Omata M., Ohto M.;
RT "Heterogeneity and evolution rates of delta virus RNA sequences.";
RL J. Virol. 64:5594-5599(1990).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
CC HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90190; BAA14214.1; -;
CC FIR; A36409; SAVILDM.
CC InterPro; IPR002506; HDV_ag.
CC Pfam; PF01517; HDV_ag; 1.
CC ProDom; PD002887; HDV_ag; 1.
KW Antigen; RNA-binding; Nuclear protein; Phosphorylation.

SQ SEQUENCE 195 AA, 21811 MW, F0APP63C16F746F1 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45

RESULT 31
AA NT HDVM2 STANDARD; PRT; 195 AA.
AC P25882;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Delta antigen.
OS Hepatitis delta virus (isolate Japanese M-2) (HDV).
OC Viruses; Deltavirus.
OX NCBI_TaxID=10425;
RN [1]
RP MEDLINE=91012805; PubMed=2214027;
RA Imazeki F., Omata M., Ohto M.;
RT "Heterogeneity and evolution rates of delta virus RNA sequences.";
RL J. Virol. 64:5594-5599(1990).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
CC HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90191; BAA14215.1; -;
CC InterPro; IPR002506; HDV_ag.
CC Pfam; PF01517; HDV_ag; 1.
CC ProDom; PD002887; HDV_ag; 1.
KW Antigen; RNA-binding; Nuclear protein; Phosphorylation.
SQ SEQUENCE 195 AA, 21799 MW, 27C708315250A442 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45

RESULT 32
VATE PYRAB STANDARD; PRT; 199 AA.
AC Q9UXU4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE V-type ATP synthase subunit E (SC 3.6.3.14) (V-type ATPase subunit E).
GN ATP6 OR PYRAB17640 OR PAB1182.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
CC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=GES / Orsay;
RA MEDLINE=22511545; PubMed=12622808;
RA Cohen G.N., Sarbe V., Plament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT Archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
CC -1- FUNCTION: Produces ATP from ADP in the presence of a proton
CC gradient across the membrane.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -1- SIMILARITY: Belongs to the V-ATPase E subunit family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AJ248288; CAB50669.1; -;
DR PIR; G75028; G75028.
DR HAMAP; MF 00311; -; 1.
DR InterPro; IPR002842; ATPsynT_Esub.
DR Pfam; PF01991; V-ATP-synt_E; 1.
DR Hydrolase; ATP synthetis; Hydrogen ion transport; Complete proteome.
KW SEQUENCE 199 AA; 23132 MW; P9230F07A89D5B6 CRC64;
SQ
Query Match 1.3%; Score 7; DB 1; Length 199;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 346 LKKEAVE 352
DB 109 LKKEAVE 115
|||||
RESULT 33
TRPF STAPP STANDARD; PRT; 207 AA.
AC QSCS5;
DT 10-OCT-2003 (Rel. 42, Created)
DF 10-OCT-2003 (Rel. 42, Last sequence update)
DE N-(5'-phosphoribosyl)anthranilate isomerase (EC 5.3.1.24) (PRAI).
GN TRPF OR SE1052.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593(2003).
CC -1- CATALYTIC ACTIVITY: N-(5-phospho-beta-D-ribose)-anthranilate = 1-
CC (2-carboxyphenylamino)-1-deoxy-D-ribose 5-phosphate.
CC -1- PATHWAY: Tryptophan biosynthesis; third step.
CC -1- SIMILARITY: Belongs to the trpF family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

CC EMBL; AR016747; AAO04649.1; -;
DR HAMAP; MF 00135; -; 1.
DR InterPro; IPR001240; PRAI.
DR Pfam; PF00697; PRAI; 1.
KW Isomerase; Tryptophan biosynthesis; Complete proteome.
SQ SEQUENCE 207 AA; 23785 MW; 7337A1DF18A4982 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 207;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKKKLE 410
DB 15 KKKKLE 21
|||||
RESULT 34
YA03 ARCFU STANDARD; PRT; 211 AA.
AC Q29259;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF1003.
GN AF1003.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049433; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickley E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glöckle A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton N.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AB001034; AAB90239.1; -;
DR PIR; C69375; C69375.
DR TIGR; AF1003; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 211 AA; 24599 MW; 50411449730812F8 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 165 TLSTPEE 171
DB 33 TLSTPEE 39
|||||
RESULT 35

AAANT_HDVAM
ID AANT_HDVAM STANDARD; PRT; 214 AA.
AC P25889;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Delta antigen.
OS Hepatitis delta virus (isolate American) (HDV).
OC Viruses; Deltavirus.
OX NCBI_TaxID=10422;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87315446; PubMed=3627276;
RA Makino S., Chang M.F., Shieh C.K., Kamahora T., Vannier D.M.,
RA Govindarajan S., Lai M.M.C.;
RT "Molecular cloning and sequencing of a human hepatitis delta (delta)
RT virus RNA";
RL Nature 329:343-346(1987).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
CC HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M28267; AA98112.1;
CC PDB; 1A92; 08-JUN-99.
CC PDB; 1BY0; 26-SEP-01.
CC InterPro; IPR002506; HDV ag.
CC Pfam; PF01517; HDV ag; 1.
CC ProDom; PD002887; HDV ag; 1.
KW Antigen; RNA-binding; Nuclear protein; Phosphorylation; 3D-structure.
SQ SEQUENCE 214 AA; 24145 MW; 6B26D2FC0B59A50 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 214;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45

RESULT 36
AAANT_HDVNA
ID AANT_HDVNA STANDARD; PRT; 214 AA.
AC P25880;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Delta antigen.
OS Hepatitis delta virus (isolate Nauru) (HDV).
OC Viruses; Deltavirus.
OX NCBI_TaxID=10426;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91020976; PubMed=2219700;
RA Chao Y.C., Chang M.F., Gust I., Lai M.M.C.;
RT "Sequence conservation and divergence of hepatitis delta virus RNA";
RE Virology 178:384-392(1990).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
CC HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M58629; AA859753.1;
CC PIR; A36212; SAVLDN.
CC InterPro; IPR002506; HDV ag.
CC Pfam; PF01517; HDV ag; 1.
CC ProDom; PD002887; HDV ag; 1.
KW Antigen; RNA-binding; Nuclear protein; Phosphorylation.
SQ SEQUENCE 214 AA; 24180 MW; 91F2D2B55EC801D CRC64;

Query Match 1.3%; Score 7; DB 1; Length 214;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45

RESULT 37
SGAH_MYCFN
ID SGAH_MYCFN STANDARD; PRT; 218 AA.
AC P75293;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable hexulose-6-phosphate synthase (BC 4.1.2.-) (HUMPS) (D-arabino
DE 3-hexulose 6-phosphate formaldehyde lyase).
GN SGAH OR MPN493 OR MP349.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=1104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=97419490; PubMed=9274005;
RA Reizer J., Reizer A., Sailer M.H. Jr.;
RT "Is the ribulose monophosphate pathway widely distributed in
RT bacteria?";
RL Microbiology 143:2519-2520(1997).
CC -!- FUNCTION: Condensation of D-ribulose 5-phosphate with formaldehyde
CC to form D-arabino-6-hexulose 3-phosphate.
CC -!- PATHWAY: Probably part of a sugar metabolic pathway along with
CC sgaU and sgaE.
CC -!- SIMILARITY: BELONGS TO THE HUMPS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000033; AA895997.1;
CC PIR; S73675; S73675.
CC InterPro; IPR001754; OMPdecase.
CC Pfam; PF00215; OMPdecase; 1.
KW Lyase; Complete proteome.

SQ SEQUENCE 218 AA; 24133 MW; 5285E0388F1CBA4 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 264 TLNDIQL 270
DB 175 TLNDIQL 181
RESULT 38
RP29 HUMAN
ID RP29 HUMAN STANDARD; PRT; 220 AA.
AC O95707; OSUOQ3;
DF 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ribonuclease P protein subunit p29 (EC 3.1.26.5) (hPOP4).
GN RPP29 OR POP4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 86-95.
RP MEDLINE=99146772; PubMed=10024167;
RA Jarrous N., Eder P.S., Wesolowski D., Altman S.;
RT "Rpp14 and Rpp29, two protein subunits of human ribonuclease P";
RL RNA 5:153-157(1999).
RN [2]
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RP MEDLINE=99289489; PubMed=10352175;
RA van Bennekom H., Pruijn G.J.M., van Venrooij W.J.;
RT "hPop4: a new protein subunit of the human RNase MRP and RNase P
ribonucleoprotein complexes";
RL Nucleic Acids Res. 27:2465-2472(1999).
RN [3]
RP SEQUENCE FROM N.A.
RP TISSUE=Lung, and Lymph;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalius D.E.,
RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
CC -1- FUNCTION: Part of ribonuclease P, a protein complex that generates
mature tRNA molecules by cleaving their 5' ends. May function with
RPP38 to coordinate the nucleolar targeting and/or assembly of
RNase P.
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage of RNA, removing 5'-
extra-nucleotide from tRNA precursor.
CC -1- SUBUNIT: RNase P consists of a RNA moiety and at least 8 protein
subunits; POP1, RPP14, RPP20/POP7, RPP25, RPP29/POP4, RPP30,
RPP38 and RPP40.
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -1- SIMILARITY: TO YEAST POP4 AND S.POMBE SPBC1703.01C.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; AF061176; AAD00893.2; -;
EXBL; Y18863; CAB39167.1; -;
DR EXBL; BC004438; AAO04438.1; -;
DR EXBL; BC006098; AAH06098.1; -;
DR GK; O95707; -;
DR MIM; 606114; -;
DR GO; GO:0000172; C:ribonuclease MRP complex; TAS.
DR GO; GO:0004526; F:ribonuclease P activity; TAS.
DR GO; GO:0003723; F:RNA binding; TAS.
DR GO; GO:0006364; P:RNA processing; TAS.
DR GO; GO:0008033; P:RNA processing; TAS.
DR InterPro; IPR002730; UPF0086.
DR Pfam; PF01868; UPF0086; 1.
DR SMART; SM00538; POP4; 1.
KW Hydrolase; Nuclear protein; tRNA processing.
FT DOMAIN 65 74 POLY-DYS.
FT CONFLICT 38 39 ST -> TS (IN REF. 1).
SQ SEQUENCE 220 AA; 25424 MW; 26141947B9F5C61E CRC64;
Query Match 1.3%; Score 7; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 146 KKKAKGL 152
DB 70 KKKAKGL 76
RESULT 39
YNCC_ECOLI
ID YNCC_ECOLI STANDARD; PRT; 221 AA.
AC P76114; P77347;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical transcriptional regulator yncC.
GN YNCC OR B1450.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Maki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampaio G., Seki Y., Sivasubramanian S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map";
RL DNA Res. 3:363-377(1996).
RN [3]
CC -1- SIMILARITY: Belongs to the gntR family of transcriptional


```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AE000242; AAC74532.1; ALT_INIT.
DR EMBL; D90784; BAA15082.1; -.
DR EMBL; D90785; BAA15085.1; -.
DR EcoGene; EG13773; yncC.
DR InterPro; IPR000524; HTH_GntR.
DR Pfam; PF003192; GntR; 1.
DR SMART; SM00345; HTH_GNTR; 1.
DR PROSITE; PS00043; HTH_GNTR_FAMILY; FALSE NEG.
KW Hypothetical protein; Transcription regulation; DNA-binding;
FT DNA BIND 37 56 H-T-H MOTIF (POTENTIAL).
ST SEQUENCE 221 AA; 25151 MW; 7A589BCFAE20D75C CRC64;
Query Match 1.3%; Score 7; DB 1; Length 221;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 324 KIQQAQ 330
DB 118 KIQQAQ 124
RESULT 40
TLIY ARATH STANDARD; PRT; 225 AA.
AC Q9SN33;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Thylakoid lumenal 17.9 kDa protein, chloroplast precursor.
GN AFG24930 OR F13K23.70.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
RP STRAIN=ev. Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller A., Stiekema W., Entian K.-D., Terryn N.,
RA Pohl T., Duesterhoeft A., Grivell L.A., Rieger M.,
RA Harris B., Ansoorge W., Brandt P., Obermaier B., Mache R., Mueller M.,
RA Weichselgartner M., de Simone V., Watson M., Schmidttheil T.,
RA Kreis M., Delsen M., Puigdomenech P., Boutry M., Bancroft I.,
RA Reichert B., Portetelle D., Perez-Alonso M., Wedler H., Ridley P.,
RA Vos P., Hohnel J., Zimmermann W., Bihlham L., Robben J.,
RA Langham S.-A., McCullagh B., Chuang Y.-J., Vandenbussche F.,
RA Van der Schueren J., Grymonprez B., Bastiaens I., Aert R., Defoor E.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Hilbert H., Braun M.,
RA Weitzenegger T., Bothe G., Ransperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Mooljman P., Klein Lankhorst R., Rose M., Haut J., Koetter P.,
RA Berreiter S., Hempel S., Feldpausch M., Lamberth S., Van den Baele H.,
RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharif M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argiriou A., Vitale D., Ligouri R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,

```

```

RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chedor P., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Baigues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielek C.,
RA Fishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Skonking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Lestraille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Grant E., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana."
RL Nature 402:769-777(1999).
RN [2]
RP SEQUENCE OF 64-93.
RC STRAIN=ev. Columbia;
RA Schubert M., Peterson U., Funk C., Schroeder W.P., Kieselbach T.;
RL Submitted (AUG-2001) to Swiss-Prot.
CC -1- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AL035523; CAB36735.1; -.
DR EMBL; AL161562; CAB79402.1; -.
DR FRL; T05514; T05514.
KW Chloroplast; Thylakoid; Transit peptide.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT TRANSIT 63 ? THYLAKOID.
FT CHAIN 64 225 THYLAKOID LUMENAL 17.9 KDA PROTEIN.
SQ SEQUENCE 225 AA; 24695 MW; 067C3FEDC1FP4C41 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 225;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 291 ERLKKLI 297
DB 205 ERLKKLI 211
RESULT 41
PYRH AERPE STANDARD; PRT; 227 AA.
ID PYRH AERPE
AC Q9YFZ0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable uridylylate kinase (EC 2.7.4.-) (UK) (Uridine monophosphate
DE kinase) (UMP kinase).
GN PYRH OR APB0401.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X1;

```



```

RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-ko K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kuchida N., Onuchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
CC -I- FUNCTION: Catalyzes the phosphorylation of UMP to UDP (by
CC similarity).
CC -I- CATALYTIC ACTIVITY: ATP + UMP = ADP + UDP
CC -I- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; first
CC step.
CC -I- SIMILARITY: Belongs to the UMP kinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AP000059; BAA79356.1; ALT_INIT.
DR HAMAP; MF 01220; -; 1.
DR InterPro; IPR001048; Aa_kinase.
DR Pfam; PF00636; aakinese; 1.
KW Transferrase; Kinase; Pyrimidine biosynthesis; Complete proteome.
SQ SEQUENCE 227 AA; 24400 MW; C1691C26A240FA6A CRC64;

Query Match 1.3%; Score 7; DB 1; Length 227;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 RQLEDIL 67
Db 167 RQLEDIL 173

RESULT 42
REGO_BP82 STANDARD; PRT; 229 AA.
AC P13870;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Antitermination protein Q.
GN Q.
OS Bacteriophage 82.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda-like viruses.
OX NCBI_TaxID=10705;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87308148; PubMed=3624233;
RA Golliger J.A., Roberts J.W.;
RT "Bacteriophage 82 gene Q and Q protein. Sequence, overproduction, and
RT activity as a transcription antiterminator in vitro."
RL J. Biol. Chem. 262:11721-11725(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96196428; PubMed=8648624;
RA Mandi A.A., Sharples G.J., Mandal T.N., Lloyd R.G.;
RT "Holiday junction resolvases encoded by homologous rusa genes in
RT Escherichia coli K-12 and phage 82."
RL J. Mol. Biol. 257:561-573(1996).
CC -I- FUNCTION: Positively regulates expression of the phage late gene
CC operons. Bacterial host RNA polymerase modified by antitermination
CC proteins transcribes through termination sites that otherwise
CC prevent expression of the regulated genes.
CC

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; J02803; AAA32298.1; -.
DR EMBL; X92588; CAA63332.1; -.
DR PIR; A29791; PQBP82.
KW Transcription regulation; Transcription termination; DNA-binding.
FT DNA BIND 14 32 POTENTIAL.
SQ SEQUENCE 225 AA; 26400 MW; E6EDABE818498667 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 ERLKKLI 297
Db 144 ERLKKLI 150

RESULT 43
MYPO_HUMAN STANDARD; PRT; 248 AA.
AC P25189; Q16072; Q92677; Q9BR67;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myelin P0 protein precursor (Myelin protein zero) (Myelin peripheral
DE protein) (MPP).
DE GN MPZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92062068; PubMed=1719967;
RA Hayasaka K., Nanao K., Tahara M., Sato W., Takada G., Miura M.,
RA Uemura K.;
RT "Isolation and sequence determination of cDNA encoding the major
RT structural protein of human peripheral myelin."
RL Biochem. Biophys. Res. Commun. 180:515-518(1991).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT CMT1B HTS-98.
RC TISSUE=Spinal cord;
RX MEDLINE=93356807; PubMed=7688964;
RA Hayasaka K., Ohnishi A., Takada G., Fukushima Y., Murai Y.;
RT "Mutation of the myelin P0 gene in Charcot-Marie-Tooth neuropathy
RT type 1."
RL Biochem. Biophys. Res. Commun. 194:1317-1322(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=94154677; PubMed=7509228;
RA Pham-Dinh D., Fourbil Y., Blanguet F., Mattei M.-G., Roeckel N.,
RA Latour P., Chazot G., Vandenbergh A., Dautigny A.;
RT "The major peripheral myelin protein zero gene: structure and
RT localization in the cluster of Fc gamma receptor genes on human
RT chromosome 1q21.3-q23."
RL Hum. Mol. Genet. 2:2051-2054(1993).
RN [4]
RP SEQUENCE FROM N.A.
RA Kalline N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,
RA Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y.,
RA Phelan M., Farmer A.;
RT "Cloning of human full-length CDSs in BD Creator(TM) system donor
RT vector."
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.

```

RC TISSUE=Skin;
EX MEDLINE=2338257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,
RA Bosak S.A., McWan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny N.J., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettunen M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.F., Skalska U., Smalley D.B.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 24-248 FROM N.A.
RA Nelis E., Timmerman V., De Jonghe P., Myllye L., Martin J.-J.,
RA Van Broeckhoven C.;
RT "Linkage and mutation analysis in an extended family with Charcot-
RT Marie-Tooth disease type 1B.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
RN [7]
RP REVIEW ON CMT1B VARIANTS.
EX MEDLINE=95282670; PubMed=7762451;
RA Roa B.B., Lupski J.R.;
RL Adv. Hum. Genet. 22:117-152(1994).
RN [8]
RP REVIEW ON CMT1B VARIANTS.
EX MEDLINE=94302675; PubMed=7518101;
RA Patel P.I., Lupski J.R.;
RT "Charcot-Marie-Tooth disease: a new paradigm for the mechanism of
RT inherited disease.";
RL Trends Genet. 10:128-133(1994).
RN [9]
RP REVIEW ON CMT1B AND DSS VARIANTS.
EX MEDLINE=99103460; PubMed=9889385;
RA Nelis E., Haates N., van Broeckhoven C.;
RT "Mutations in the peripheral myelin genes and associated genes in
RT inherited peripheral neuropathies.";
RL Hum. Mutat. 13:11-28(1999).
RN [10]
RP VARIANT CMT1B MET-30.
EX MEDLINE=94061030; PubMed=7694726;
RA Hayasaka K., Takada G., Ionescu V.V.;
RT "Mutation of the myelin P0 gene in Charcot-Marie-Tooth neuropathy
RT type 1B.";
RL Hum. Mol. Genet. 2:1369-1372(1993).
RN [11]
RP VARIANT CMT1B CY5-82.
EX MEDLINE=94083941; PubMed=7505151;
RA Himoro M., Yoshikawa H., Matsui T., Mitsui Y., Takahashi M.,
RA Kado M., Nishimura T., Sawashita Y., Takada G., Hayasaka K.;
RT "New mutation of the myelin P0 gene in a pedigree of
RT Charcot-Marie-Tooth neuropathy 1.";
RL Biochem. Mol. Biol. Int. 31:169-173(1993).
RN [12]
RP VARIANTS CMT1B GLU-90 AND GLU-96.
EX MEDLINE=94035113; PubMed=7693129;
RA Hayasaka K., Himoro M., Sato W., Takada G., Ueyemura K., Shimizu N.,
RA Bird T.D., Conneally P.M., Chance P.F.;
RT "Charcot-Marie-Tooth neuropathy type 1B is associated with mutations
RT of the myelin P0 gene.";
RL Nat. Genet. 5:31-34(1993).
RN [13]

RP VARIANT CMT1B SER-63 DEL.
EX MEDLINE=94035114; PubMed=7693130;
RA Kalkens T., Bolhuis P.A., Wolterman R.A., Kemp S., Te Nijenhuis S.,
RA Valentijn L.J., Hensels G.W., Jemsekens F.G., de Visser M.,
RA Hoogendijk J.E., Baas F.;
RT "Deletion of the serine 34 codon from the major peripheral myelin
RT protein P0 gene in Charcot-Marie-Tooth disease type 1B.";
RL Nat. Genet. 5:35-39(1993).
RN [14]
RP VARIANT CMT1B GLU-96.
EX MEDLINE=94068501; PubMed=7504284;
RA Su Y., Brooks D.G., Li L., Lepercq J., Trofatter J.A., Ravetch J.V.,
RA Lebo R.V.;
RT "Myelin protein zero gene mutated in Charcot-Marie-Tooth type 1B
RT patients.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10856-10860(1993).
RN [15]
RP VARIANTS DSS CY5-63 AND ARG-167.
EX MEDLINE=94100981; PubMed=7506095;
RA Hayasaka K., Himoro M., Sawashita Y., Manao K., Takahashi T.,
RA Takada G., Nicholson G.A., Ouyrier R.A., Tachi N.;
RT "De novo mutation of the myelin P0 gene in Dejerine-Sottas disease
RT (hereditary motor and sensory neuropathy type III).";
RL Nat. Genet. 5:266-268(1993).
RN [16]
RP VARIANTS CMT1B LEU-78 AND ASN-134.
EX MEDLINE=95080774; PubMed=7527371;
RA Nelis E., Timmerman V., de Jonghe P., Vandenbergh A., Pham-Dinh D.,
RA Dautigny A., Martin J.-J., van Broeckhoven C.;
RT "Rapid screening of myelin genes in CMT1 patients by SSCP analysis:
RT identification of new mutations and polymorphisms in the P0 gene.";
RL Hum. Genet. 94:653-657(1994).
RN [17]
RP VARIANT CMT1B PHE-63.
EX MEDLINE=96432254; PubMed=8835320;
RA Blanquet-Grossard P., Pham-Dinh D., Dautigny A., Latour P.,
RA Bonnebouche C., Corbillion E., Chazot G., Vandenbergh A.;
RT "Charcot-Marie-Tooth type 1B neuropathy: third mutation of serine 63
RT codon in the major peripheral myelin glycoprotein P0 gene.";
RL Clin. Genet. 48:281-283(1995).
RN [18]
RP VARIANTS CMT1B LEU-78 AND CY5-101.
EX MEDLINE=96055517; PubMed=7550231;
RA Latour P., Blanquet F., Nelis E., Bonnebouche C., Chapon F.,
RA Diraison P., Ollagnon E., Dautigny A., Pham-Dinh D., Chazot G.,
RA Boucherat M., van Broeckhoven C., Vandenbergh A.;
RT "Mutations in the myelin protein zero gene associated with
RT Charcot-Marie-Tooth disease type 1B.";
RL Hum. Mutat. 6:50-54(1995).
RN [19]
RP VARIANT DSS PHE-64 DEL.
EX MEDLINE=96212920; PubMed=8630052;
RA Ikegami T., Nicholson G.A., Ikeda H., Ishida A., Johnston H., Wise G.,
RA Ouyrier R.A., Hayasaka K.;
RT "A novel homozygous mutation of the myelin P0 gene producing
RT Dejerine-Sottas disease (hereditary motor and sensory neuropathy
RT type III).";
RL Biochem. Biophys. Res. Commun. 222:107-110(1996).
RN [20]
RP VARIANTS CMT1B THR-135 AND SER-137.
EX MEDLINE=96253736; PubMed=8664899;
RA Roa B.B., Warner L.E., Garcia C.A., Russo D., Lovelace R.,
RA Chance P.F., Lupski J.R.;
RT "Myelin protein zero (MPZ) gene mutations in nonduplication type 1
RT Charcot-Marie-Tooth disease.";
RL Hum. Mutat. 7:36-45(1996).
RN [21]
RP VARIANT CMT1B SER-122.
EX MEDLINE=97001227; PubMed=8844219;
RA Blanquet-Grossard P., Pham-Dinh D., Dautigny A., Latour P.,
RA Bonnebouche C., Diraison P., Chapon F., Chazot G., Vandenbergh A.;
RT "Charcot-Marie-Tooth type 1B neuropathy: a mutation at the single
RT glycosylation site in the major peripheral myelin glycoprotein P0.";

AC Q57083;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE HTH-type transcriptional regulator perR (Peroxide resistance protein
DE perR).
GN PER OR B0254.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Loewen P.C., Swisala J., Volkert M.R.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner P.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474[1997].
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sampaio G., Mizobuchi K.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the
RT 4.0 - 6.0 min (189,987 - 281,416bp) region.";
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Laskhari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: APPARENT REGULATORY GENE INVOLVED IN PEROXIDE RESISTANCE
CC IN STATIONARY PHASE.
CC -!- SIMILARITY: Contains 1 HTH lyseR-type DNA-binding domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U57080; AAC01994.1; -;
CC EMBL; AE000133; AAC73357.1; -;
CC EMBL; D83336; BAA77924.1; -;
CC EMBL; U70214; AAC08673.1; -;
CC PIR; F64750; F64750.
CC EcoGene; EG13340; perR.
CC InterPro; IPR000847; HTH LysR.
CC InterPro; IPR005119; LysR_subst.
CC Pfam; PF00126; HTH_1; 1
CC Pfam; PF03466; LysR_substrate; 1.
CC PRINTS; PRO0039; HTHLYSR.
CC PROSITE; PS0931; HTH LYSR; 1.
CC Transcription regulation; DNA-binding; Complete proteome.
FT DOMAIN 7 64 HTH LYSR-TYPE.
FT DNA_BIND 24 44 H-T-H MOTIF (POTENTIAL).
SQ SEQUENCE 297 AA; 33638 MW; E197CE2334B7420E CRC64;
Query Match 1.3%; Score 7; DB 1; Length 297;
Best Local Similarity 100.0%; Pred. No. 73;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 289 LAERLKK 295
Db 173 LAERLKK 179
RESULT 46
MURB CLOPE STANDARD; PRT; 304 AA.
AC Q8XNIO;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosaminylglucosamine reductase (EC 1.1.1.158) (UDP-N-
DE acetylglucosamine dehydrogenase).
GN MURB OR CPE0353.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=1192842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC -!- FUNCTION: Cell wall formation (by similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylmuramate + NADP(+) = UDP-N-
CC acetyl-3-O-(1-carboxyvinyl)-D-glucosamine + NADPH.
CC -!- COFACTOR: PAD (by similarity).
CC -!- PATHWAY: Peptidoglycan biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Belongs to the murB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003186; BAB80059.1; -;
CC HAMAP; MF 00037; -; 1.
CC InterPro; IPR003170; MurB
CC InterPro; IPR006094; Oxid_PAD bind.
CC Pfam; PF01585; PAD_binding_4; 1.
CC Pfam; PF02873; MurB_C; 1.
CC Oxidoreductase; NADP; Flavoprotein; PAD; Cell wall; Cell division;
CC Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 304 AA; 33267 MW; CB4C78C107A6694B CRC64;
Query Match 1.3%; Score 7; DB 1; Length 304;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 46 TAQSGAL 52
Db 101 TAQSGAL 107
RESULT 47
Y280 ARATH STANDARD; PRT; 308 AA.
ID Y280 ARATH
AC Q82368;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein At2g29880.

GN AT2G29880 OR F6K5.1
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=106171797;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gil J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Niernman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768 (1999).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC007113; AB23628.1;
DR EMBL; AC005496; AA015055.1;
DR PIR; G84701; G84701.
KW Hypothetical protein.
SQ SEQUENCE 308 AA; 35622 MW; 9CBD18E273C09BES CRC64;

Query Match 1.3%; Score 7; DB 1; Length 308;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 336 EERHORE 342
Db 242 EERHORE 248

RESULT 48
ID HTPX THEAC STANDARD; PRT; 317 AA.
AC Q9HJV2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable protease htpx homolog (EC 3.4.24.-).
GN HTPX OR TA0861.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum.";
RL Nature 407:508-513 (2000).
CC -!- COPACITOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -!- SIMILARITY: Belongs to peptidase family M48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL445065; CAC11990.1;
DR MEROPS; M48; UPB; --;
DR HAMAP; MF 00188; --; 1.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001915; Peptidase_M48.
DR Pfam; PF01435; Peptidase_M48; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW Transmembrane; Hydrolase; Metalloprotease; Zinc; Complete proteome.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 41 61 POTENTIAL.
FT TRANSMEM 158 178 POTENTIAL.
FT TRANSMEM 189 209 POTENTIAL.
FT METAL 146 146 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 147 147 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 150 150 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 317 AA; 35616 MW; 535EFBA7F733CDEE CRC64;

Query Match 1.3%; Score 7; DB 1; Length 317;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 448 IORLEKL 454
Db 308 IORLEKL 314

RESULT 49
ID THIL METJA STANDARD; PRT; 319 AA.
AC Q60337;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable thiamine-monophosphate kinase (EC 2.7.4.16) (Thiamine-
DE phosphate kinase).
GN THIL OR MJ0028.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Clayton R.A., Gocayne J.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Tomb J.-F., Adams M.D., Reich C.I.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Weissman J.M., Glöckle A.,
RA Overbeek R., Kirkness B.F., Weinstock K.G., Merrick J.L., Nguyen D.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Gilek A.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1053-1073 (1996).
CC -!- CATALYTIC ACTIVITY: ATP + thiamine phosphate = ADP + thiamine
CC diphosphate.
CC -!- PATHWAY: Thiamine biosynthesis.
CC -!- SIMILARITY: Belongs to the thiamine-monophosphate kinase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

```
CC EMBL; U67461; AAB98008.1; -.
DR F1R; D64303; D64303.
DR TIGR; MJO028; -.
DR InterPro; IPR000728; AIR synth.
DR InterPro; IPR006283; ThiL.
DR Pfam; PF00586; AIRS; 1.
DR TIGRfams; TIGR01379; ThiL; 1.
DR Thiamine biosynthesis; transferase; Kinase; Complete proteome.
KW SEQUENCE 319 AA; 36160 MW; 92607DSAD59A723F CRC64;
SQ
Query Match 1.3%; Score 7; DB 1; Length 319;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 151 GLGKEIT 157
DB 222 GLGKEIT 228
RESULT 50
RPOA_UREPA
ID RPOA_UREPA STANDARD; PRT; 327 AA.
AC Q9PQ04;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (RNAP alpha subunit) (Transcriptase alpha chain) (RNA polymerase alpha subunit).
GN RPOA OR U0257.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Setovar 3; PubMed=11048724;
RX MEDLINE=20500219;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y., Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma urealyticum."
RL Nature 407:757-762(2000).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + [RNA] (N).
CC -!- SUBUNIT: Homodimer. The RNAP catalytic core consists of 2 alpha, 1 beta, 1 beta', and 1 omega subunit (By similarity).
CC -!- DOMAIN: The N-terminal domain is essential for RNAP assembly and basal transcription, whereas the C-terminal domain is involved in interaction with transcriptional regulators and with upstream promoter elements (By similarity).
CC -!- SIMILARITY: Belongs to the RNA polymerase alpha chain family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements/ or send an email to license@ebi.ac.uk).
```

```
KW Complete proteome.
FT DOMAIN 1 242 ALPHA N-TERMINAL DOMAIN (ALPHA-NTD)
FT (BY SIMILARITY).
FT DOMAIN 259 327 ALPHA C-TERMINAL DOMAIN (ALPHA-CTD)
FT (BY SIMILARITY).
SQ SEQUENCE 327 AA; 36817 MW; 4BAEC583237902E4 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 327;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 248 REEEKR 254
DB 254 REEEKR 260
RESULT 51
IF2B_MOUSE
ID IF2B_MOUSE STANDARD; PRT; 331 AA.
AC Q9SL45; O9CSH6; O9CT12;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Eukaryotic translation initiation factor 2 subunit 2 (Eukaryotic translation initiation factor 2 beta subunit) (eIF-2-beta).
GN EIF2S2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T., Baldairelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schirler L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S., Dalla B., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Guetincich S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming I.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D., Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K., Hopkins R.F., Jordan H., Moore F., Max S.I., Wang J., Hoie L., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Donaldo M.F., Casavant T.L., Scheetz T.E.,
```


RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abranson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnurra A., Schein J.E., Jones S.J.M., Marra M.A.,
RT human and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: eIF-2 functions in the early steps of protein synthesis
CC by forming a ternary complex with GTP and initiator tRNA. This
CC complex binds to a 40S ribosomal subunit, followed by mRNA binding
CC to form a 43S preinitiation complex. Junction of the 60S ribosomal
CC subunit to form the 80S initiation complex is preceded by
CC hydrolysis of the GTP bound to eIF-2 and release of an eIF-2-GDP
CC binary complex. In order for eIF-2 to recycle and catalyze another
CC round of initiation, the GDP bound to eIF-2 must exchange with GTP
CC by way of a reaction catalyzed by eIF-2b (3y similarity).
CC -!- SUBUNIT: Heterotrimer composed of an alpha, a beta and a gamma
CC chain (3y similarity).
CC -!- SIMILARITY: Belongs to the eIF-2-beta / eIF-5 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK011503; BAB27663.1; -;
DR EMBL; AK0112817; BAB28490.2; ALT_INIT.
DR EMBL; BC003848; AAB03848.1; -;
DR MGDB; MG1:1914454; Eif2s2.
DR InterPro; IPR002735; eIF5_eIF2B.
DR Pfam; PF01873; eIF5_eIF2B; 1.
DR ProDom; PD004078; eIF5_eIF2B; 1.
DR SMART; SM00653; eIF2B 5; 1.
KW Initiation factor; Protein biosynthesis; Zinc-finger.
FT DOMAIN 14 21 POLY-LYS (BASIC).
FT DOMAIN 79 87 POLY-LYS (BASIC).
FT DOMAIN 124 129 POLY-LYS (BASIC).
FT ZN_FING 279 303 C4-TYPE (POTENTIAL).
SQ SEQUENCE 331 AA; 38092 MW; E91BEBBF330B7A5C CRC64;
Query Match 1.3%; Score 7; DB 1; Length 331;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 87 EPAEPED 93
Db 109 EPAEPED 115
|||||
RESULT 52
IF2B_RABIT STANDARD; PRT; 333 AA.
AC P41035;
ID IF2B_RABIT
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Eukaryotic translation initiation factor 2 subunit 2 (Eukaryotic
DE translation initiation factor 2 beta subunit) (eIF-2-beta).
GN EIF2B2 OR EIF2B.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;

RN SEQUENCE FROM N.A.
RC STRAIN=New Zealand white.
RX MEDLINE=94032484; PubMed=8218412;
RA Price N.T., Hall L., Proud C.G.;
RT "Cloning of cDNA for the beta-subunit of rabbit translation initiation
RL factor-2 using PCR.";
RN Biochim. Biophys. Acta 1216:170-172(1993).
RP [2]
RP SEQUENCE OF 25-39; 122-136 AND 195-218.
RC TISSUE=Reticulocytes;
RX MEDLINE=92002150; PubMed=1911855;
RA Bommer U.-A., Kraft R., Kurzchalia T.V., Price N.T., Proud C.G.;
RT "Amino acid sequence analysis of the beta- and gamma-subunits of
RL eukaryotic initiation factor eIF-2. Identification of regions
RN interacting with GTP.";
RP Biochim. Biophys. Acta 1079:308-315(1991).
RC [3]
RC CHARACTERIZATION.
RX TISSUE=Reticulocytes;
RX MEDLINE=87240215; PubMed=3592677;
RA Schafer M.P., Fairwell T., Parker D.S., Knight M., Anderson W.F.,
RT "The purification and characterization of subunits alpha, beta, and
RL gamma from the rabbit reticulocyte eukaryotic initiation factor 2.";
CC Arch. Biochem. Biophys. 255:337-346(1987).
CC -!- FUNCTION: eIF-2 functions in the early steps of protein synthesis
CC by forming a ternary complex with GTP and initiator tRNA. This
CC complex binds to a 40S ribosomal subunit, followed by mRNA binding
CC to form a 43S preinitiation complex. Junction of the 60S ribosomal
CC subunit to form the 80S initiation complex is preceded by
CC hydrolysis of the GTP bound to eIF-2 and release of an eIF-2-GDP
CC binary complex. In order for eIF-2 to recycle and catalyze another
CC round of initiation, the GDP bound to eIF-2 must exchange with GTP
CC by way of a reaction catalyzed by eIF-2b.
CC -!- SUBUNIT: Heterotrimer composed of an alpha, a beta and a gamma
CC chain.
CC -!- PTM: The N-terminus is blocked.
CC -!- SIMILARITY: Belongs to the eIF-2-beta / eIF-5 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73836; CAA52058.1; -;
DR PIR; JC2329; JC2329.
DR PIR; S17871; S17871.
DR InterPro; IPR002735; eIF5_eIF2B.
DR Pfam; PF01873; eIF5_eIF2B; 1.
DR ProDom; PD004078; eIF5_eIF2B; 1.
DR SMART; SM00653; eIF2B 5; 1.
KW Initiation factor; Protein biosynthesis; Zinc-finger.
FT DOMAIN 14 21 POLY-LYS (BASIC).
FT DOMAIN 79 87 POLY-LYS (BASIC).
FT DOMAIN 124 129 POLY-LYS (BASIC).
FT ZN_FING 281 305 C4-TYPE (POTENTIAL).
SQ SEQUENCE 333 AA; 38326 MW; 3FFDBAB92DFC1465 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 87 EPAEPED 93
Db 109 EPAEPED 115
|||||
RESULT 53
SYFA_CLONE

ID SYFA CLOAB STANDARD; PRT; 339 AA.
AC Q97GK9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phenylalanyl-tRNA synthetase alpha chain (EC 6.1.1.20)
DE (Phenylalanyl-tRNA ligase alpha chain) (PHERS).
GN PHERS OR CAC2357.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=2135925; PubMed=11466286;
RA Noelling J., Berton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 193:4823-4838 (2001).
CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC diphosphate + L-phenylalanyl-tRNA(Phe).
CC -!- COPACTOR: Binds 2 magnesium ions per tetramer (By similarity).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC Phe-tRNA synthetase alpha chain subfamily 1.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AE007736; AAK80313.1; -
DR PIR; F97190; F97190.
DR HAVAP; MF_00281; -; 1.
DR InterPro; IPR004188; Phe_tRNA_synt_N.
DR InterPro; IPR004529; PheS.
DR InterPro; IPR002319; tRNA-synt 2d.
DR InterPro; IPR006195; tRNA-synt N; I.
DR Pfam; PF02912; Phe_tRNA_synt_N; I.
DR Pfam; PF01409; tRNA-synt 2d; I.
DR TIGRFAMs; TIGR00468; pheS; I.
DR PROSITE; PSS0862; AA_tRNA_LIGASE II; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Metal-binding; Magnesium; Complete proteome.
FT METAL 254 254 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 339 AA; 38508 MW; 229E76449BBE10918 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 339;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 222 RSKLES 228
| | | | |
Db 65 RSKLES 71

RESULT 54
ID IAP_ECOLI
ID IAP_ECOLI STANDARD; PRT; 345 AA.
AC P10423;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alkaline phosphatase isozyme conversion protein precursor

(EC 3.4.11.1-).
DE IAP OR B2753.
GN Escherichia coli.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=88058748; PubMed=3316184;
RX Ishino Y., Shinagawa H., Makino K., Anemura M., Nakata A.;
RT "Nucleotide sequence of the iap gene, responsible for alkaline
phosphatase isozyme conversion in Escherichia coli, and
RT identification of the gene product.";
RL J. Bacteriol. 169:5429-5433 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna M.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [3]
RP SEQUENCE OF 1-111 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92268080; PubMed=1316900;
RA Leyh T.S., Vogt T.F., Suo Y.;
RT "The DNA sequence of the sulfate activation locus from Escherichia
RT coli K-12.";
RL J. Biol. Chem. 267:10405-10410 (1992).
CC -!- FUNCTION: This protein, presumably an aminopeptidase, mediates the
CC conversion of E.coli alkaline phosphatase isozyme 1, to isozymes 2
CC and 3 by removing, one by one, the two amino-terminal arginine
CC residues.
CC -!- SIMILARITY: Belongs to peptidase family M28C.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U18270; AAK24005.1; -
DR EMBL; U29579; AAK69283.1; -
DR EMBL; AE000359; AAC75795.1; -
DR EMBL; M74586; AAA23644.1; -
DR PIR; A28382; KZEC.
DR MEROPS; M28.005; -
DR EcoGene; EG10488; iap.
DR InterPro; IPR007484; Peptidase_M28.
DR Pfam; PF04389; Peptidase_M28; I.
KW Hydrolase; Aminopeptidase; Signal; Complete proteome.
FT SIGNAL 1 24 PROBABLE.
FT CHAIN 25 345 ALKALINE PHOSPHATASE ISOZYME CONVERSION
FT PROTEIN.
SQ SEQUENCE 345 AA; 37920 MW; 3DC433FB569E912D CRC64;

Query Match 1.3%; Score 7; DB 1; Length 345;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 288 ELAERLK 294
| | | | |
Db 153 ELAERLK 159

RESULT 55
ID RFI_TREPA STANDARD; PRT; 351 AA.
ID RFI_TREPA

```

AC 083090;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptide chain release factor 1 (RF-1).
GN PRFA OR TP0051.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalai H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artisch P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT Spirochete."
RL Science 281:375-388(1998).
CC -!- FUNCTION: Peptide chain release factor 1 directs the termination
CC of translation in response to the peptide chain termination codons
CC UAG and UAA (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the prokaryotic/mitochondrial release
CC factor family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AE001190; AAC65047.1; -.
DR F1R; E71372; E71372.
DR TIGR; TP0051; -.
DR HAMAP; MF_00093; -.
DR InterPro; IPR005139; PCRF.
DR InterPro; IPR000352; Pep_rel_factor_1.
DR InterPro; IPR004373; PrfA.
DR Pfam; PF03462; PCRF; 1.
DR Pfam; PF00472; RF-1; 1.
DR TIGRFAMs; TIGR00019; prfA; 1.
DR PROSITE; PS00745; RF_PROK_1; 1.
DR Protein biosynthesis; Complete proteome.
DR SEQUENCE 351 AA; 40194 MW; 693C6F883B5797C CRC64;
SQ
Query Match 1.3%; Score 7; DB 1; Length 351;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 512 PSTEASG 518
Db 187 PSTEASG 193
|||||
RESULT 56
RF1_BACHD STANDARD; PRT; 356 AA.
AC Q9X6F4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Peptide chain release factor 1 (RF-1).
GN PRFA OR BH3775.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=205112582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji P., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- FUNCTION: Peptide chain release factor 1 directs the termination
CC of translation in response to the peptide chain termination codons
CC UAG and UAA (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the prokaryotic/mitochondrial release
CC factor family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AP001519; BAB07494.1; -.
DR F1R; G84121; G84121.
DR HAMAP; MF_00093; -.
DR InterPro; IPR005139; PCRF.
DR InterPro; IPR000352; Pep_rel_factor_1.
DR InterPro; IPR004373; PrfA.
DR Pfam; PF03462; PCRF; 1.
DR Pfam; PF00472; RF-1; 1.
DR TIGRFAMs; TIGR00019; prfA; 1.
DR PROSITE; PS00745; RF_PROK_1; 1.
DR Protein biosynthesis; Complete proteome.
DR SEQUENCE 356 AA; 40597 MW; 34342F09CBBAA8A8 CRC64;
SQ
Query Match 1.3%; Score 7; DB 1; Length 356;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 ELMKQOE 364
Db 349 ELMKQOE 355
|||||
RESULT 57
MKK2_DROME STANDARD; PRT; 359 AA.
AC P49071; Q9W480;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE MAP kinase-activated protein kinase 2 (EC 2.7.1.-) (MAPKAP-2)
DE protein kinase 2 (MAPKAP kinase 2) (MAPKAP-2).
GN MAPK-AK2 OR CG3086.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BIC-D(R26); TISSUE=Ovary;
RX MEDLINE=96011635; PubMed=7590268;
RA Laroche S., Suter B.;
RT "The Drosophila melanogaster homolog of the mammalian MAPK-activated
RT protein kinase-2 (MAPKAP-2) lacks a proline-rich N-terminus."
RL Gene 163:209-214(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;

```

MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.M.,
RA Sutton G.G., Wotman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Chape M., Pfeiffer B.D.,
Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.P., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benson P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Bottler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Kays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreria S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.C., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M., P.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert X., Remington K., Saunders R.P.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).

CC -|- FUNCTION: ITS PHYSIOLOGICAL SUBSTRATE SEEMS TO BE THE SMALL HEAT
CC -|- SHOCK PROTEIN (HSP27/HSP25) (BY SIMILARITY).
CC -|- PM: Phosphorylated and activated by MAP kinase (By similarity).
CC -|- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U20757; AAA86885.1; --
DR EMBL; AE003435; AAC22408.1; --
DR PIR; JC4297; JC4297.
DR HSP; Q63450; IA06.
DR FlyBase; F58N0013987; MAPK-AK2.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR008271; Ser Thr pkin AS.
DR InterPro; IPR002290; Ser Thr pkinase.
DR InterPro; IPR001245; Tyr pkinase.
DR Pfam; PF00069; pkinase; 1
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
KW Transferrase: Serine/threonine-protein kinase; ATP-binding;
KW Phosphorylation.

DR DOMAIN 20 281 PROTEIN KINASE.
DR NP_BIND 26 34 ATP (BY SIMILARITY).
DR BINDING 49 49 ATP (BY SIMILARITY).

```

FT  ACT SITE      142      142      BY SIMILARITY.
SQ  SEQUENCE      359 AA;  41401 MW;  69CSF9A94D511EC CRC64;

Query Match      1.3%;  Score 7;  DB 1;  Length 359;
Best Local Similarity 100.0%;  Pred. No. 86;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      331 MLKEAE 337
DB      297 MLKEAE 303

RESULT 58
YH6 YEAST
ID  _YH6 YEAST      STANDARD;      PRT;      374 AA.
AC  P38757;
DT  01-FEB-1995 (Rel. 31, Created)
DT  01-FEB-1995 (Rel. 31, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  Hypothetical 41.2 kDa protein in ERG7-NMD2 intergenic region.
GN  YHR076W.
OS  Saccharomyces cerevisiae (Baker's yeast).
OC  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC  Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX  NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
RC  STRAIN=S288C / AB972;
RX  MEDLINE=94378003; PubMed=8091229;
RA  Johnston M.; Andrews S.; Brinkman R.; Cooper J.; Ding H.; Dover J.,
RA  Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA  Kucaba T., Hillier L.W., Jier M., Johnston L., Langston S., Mouser L.,
RA  Latreille P., Louis E.J., Macri C., Mardis E., Menzies S., Nhan M.,
RA  Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA  Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA  Vaudin M.;
RT  "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT  VII."
RL  Science 265:2077-2082 (1994).
CC  1- SIMILARITY: Contains 1 PP2C-like domain.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch)
-----
CC  EMBL; U10556; A286808.1; --
CC  F1R; S46810; S46810.
CC  Germonline; 139393; --
CC  SGD; S0001118; YHR076W.
CC  GO; GO:0005739; C:mitochondrion; IDA.
CC  GO; GO:0015071; F:protein phosphatase type 2C activity; IDA.
CC  InterPro; IPR001932; PP2C-like.
CC  SMART; SM00331; PP2C_SIG; 1.
CC  SMART; SM00332; PP2C; 1.
CC  Hypothetical protein; Transmembrane.
KW  POTENTIAL.
FT  TRANSMEM      19      39
FT  DOMAIN      117      374
SQ  SEQUENCE      374 AA;  41130 MW;  74E7F084ABD157FE CRC64;

Query Match      1.3%;  Score 7;  DB 1;  Length 374;
Best Local Similarity 100.0%;  Pred. No. 89;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      330 EMLKEAE 336
DB      258 EMLKEAE 264

RESULT 59

```

```

VE2_HPV57
ID_VB2_HPV57 STANDARD; PRT; 383 AA.
AC P22155;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Regulatory protein E2.
GN E2
OS Human papillomavirus type 57.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OC NCBI_TaxID=10597;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=9118699; PubMed=1964523;
RA Hirsch-Behnam A., Delius H., de Villiers E.M.;
RT "A comparative sequence analysis of two human papillomavirus (HPV)
RL types 2a and 57."
CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -!- SUBUNIT: Binds DNA as a dimer.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5059;
RN (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94086131; PubMed=8262652;
RA Ramesh M.V., Sirakova T., Kolatukudy P.B.;
RT "Isolation, characterization, and cloning of cDNA and the gene for an
RT elastinolytic serine proteinase from Aspergillus flavus.";
RL Infect. Immun. 62:79-85(1994).
CC -!- FUNCTION: Elastolytic enzyme probably acting as a virulence factor
CC in invasive aspergillosis.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity,
CC and of BZ-Arg-OET > Ac-Tyr-OET. Does not hydrolyze peptide amides.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.,
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphohikimate =
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphohikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the EPSP synthase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE013066; AAM24270.1; -
CC HAMAP; MF_00210; -; 1.
CC InterPro; IPR006264; AroA.
CC InterPro; IPR001986; EPSP synth.
CC Pfam; PF00275; EPSP synthase; 1.
CC ProDom; PD001867; EPSP synthase; 1.
CC TIGRfams; TIGR01356; aroA; 1.
CC PROSITE; PS00104; EPSP SYNTHASE 1; 1.
CC PROSITE; PS00885; EPSP SYNTHASE 2; 1.
KW Aromatic amino acid biosynthesis; Transferrase; Complete proteome.
SQ SEQUENCE 423 AA; 46069 MW; C8BECDC1902DF51 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 423;
Best Local Similarity 100.0%; Pred. No. 1e-02; 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 385 LSKSSEV 391
DB 79 LSKSSEV 85
|||||||

RESULT 62
Y12A_ECOLI
ID Y12A_ECOLI STANDARD; PRT; 424 AA.
AC P51026; Q47535;
DT 01-OCT-1996 (Rel. 34; Created)
DT 15-JUL-1999 (Rel. 38; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Insertion element IS2A hypothetical 48.2 kDa protein.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Nashimoto H., Saito N.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the transposase family 8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D85613; BAA12836.1; -
CC PIR; S78610; S78610.
CC InterPro; IPR001584; Rve.
CC InterPro; IPR002514; Transposase_8.

DR Pfam; PF00665; rve; 1.
DR Pfam; PF01527; Transposase_8; 1.
KW Transposable element; Hypothetical protein.
SQ SEQUENCE 424 AA; 48153 MW; A791CE85D05A329 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 424;
Best Local Similarity 100.0%; Pred. No. 1e-02; 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 346 LKKEAVE 352
DB 111 LKKEAVE 117
|||||||

RESULT 63
FIBG_PETMA
ID FIBG_PETMA STANDARD; PRT; 432 AA.
AC P04115;
DT 01-NOV-1986 (Rel. 03; Created)
DT 01-NOV-1986 (Rel. 03; Last sequence update)
DT 10-OCT-2003 (Rel. 42; Last annotation update)
DE Fibrinogen gamma chain precursor.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8519776; PubMed=2581603;
RA Strong D.D., Moore M., Cottrell B.A., Bohonus V.L., Pontes M.,
RA Evans B., Riley K., Doolittle R.F.;
RT "Lamprey fibrinogen gamma chain: cloning, cDNA sequencing, and
RT general characterization.";
RL Biochemistry 24:92-101(1985).
CC -1- FUNCTION: Fibrinogen has a double function: yielding monomers that
CC polymerize into fibrin and acting as a cofactor in platelet
CC aggregation.
CC -1- SUBUNIT: HEXAMER CONTAINING 2 SETS OF 3 NONIDENTICAL CHAINS
CC (ALPHA, BETA AND GAMMA), LINKED TO EACH OTHER BY DISULFIDE BONDS.
CC -1- PTM: Conversion of fibrinogen to fibrin is triggered by thrombin,
CC which cleaves fibrinopeptides A and B from alpha and beta chains,
CC and thus exposes the N-terminal polymerization sites responsible
CC for the formation of the soft clot. The soft clot is converted
CC into the hard clot by factor XIIIa which catalyzes the epsilon-
CC (gamma-glutamyl)lysine cross-linking between gamma chains
CC (stronger) and between alpha chains (weaker) of different
CC monomers.
CC -1- SIMILARITY: Contains 1 fibrinogen C-terminal domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; K03049; AAA49262.1; -
CC PIR; A03129; FGLMGS.
CC PDB; 1LMU; 23-AUG-02.
CC PDB; 1N73; 07-JAN-03.
CC InterPro; IPR002181; Fibrinogen_C.
CC Pfam; PF00147; fibrinogen_C; 1.
CC SMART; SM00186; FBG; 1.
CC PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; 1.
KW Blood coagulation; Plasma; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 24 BY SIMILARITY.
FT CHAIN 25 432 FIBRINOGEN GAMMA CHAIN.
FT DISULFID 32 32 INTERCHAIN (WITH GAMMA CHAIN)
FT (BY SIMILARITY).
FT DISULFID 42 42 INTERCHAIN (WITH BETA CHAIN)
FT (BY SIMILARITY).
FT DISULFID 46 46 INTERCHAIN (WITH ALPHA CHAIN)

```
FT DISULFID 158 158 (BY SIMILARITY).
FT INTERCHAIN (WITH BETA CHAIN)
FT (BY SIMILARITY).
FT DISULFID 162 162 INTERCHAIN (WITH GAMMA CHAIN)
FT (BY SIMILARITY).
FT DISULFID 178 207 BY SIMILARITY.
FT DISULFID 348 361 BY SIMILARITY.
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .).
SQ SEQUENCE 432 AA; 49203 MW; B503979B296DFB24 CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 432;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 316 LQQQLVD 322
DB 147 LQQQLVD 153
|||||

RESULT 64
SDHD_FUSNN
ID SDHD_FUSNN STANDARD; PRT; 441 AA.
AC Q8RFK6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Probable D-serine dehydratase (EC 4.3.1.18) (D-serine deaminase)
DE (DSD).
DN DSDA OR FN0553.
OS Bacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium
OX NCBI_TaxID=76956;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haseel Korn R.,
RA Feinstein M., Kyriades N., Overbeek R.
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -!- CATALYTIC ACTIVITY: D-serine = pyruvate + NH(3).
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC -!- SIMILARITY: Belongs to the serine/threonine dehydratase family.
CC DsdA subfamily.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; AE010566; AAL94749.1; -
DR HAMAP; MF_01030; -
DR InterPro; IPR001926; S6 enzyme_beta.
DR InterPro; IPR000634; S/T dehydrtse_BS.
DR Pfam; PF00291; PALP; 1.
DR PROSITE; PS00165; DEHYDRATASE_SRR_THR; FALSE NEG.
KW Lyase; Pyridoxal phosphate; Complete proteome.
FT BINDING 115 115 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 441 AA; 49954 MW; E2058AF21B4851E2 CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 441;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 LKEAER 338
```

```
|||||
46 LKEAER 52

Db
RESULT 65
GAS8_HUMAN
ID GAS8_HUMAN STANDARD; PRT; 478 AA.
AC O95995;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Growth-arrest-specific protein 8 (Growth arrest-specific 11).
DE GAS8 OR GAS11.
GN Homo sapiens (Human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99009317; PubMed=9790751;
RA Whitmore S.A., Sattasatian C., Crawford J., Lower K.M., McCallum B.,
RA Seshadri R., Cornelisse C.J., Moerland E.W., Cleton-Jansen A.-M.,
RA Tipping A.J., Mathew C.G., Savino M., Savoia A., Verlander P.,
RA Auerbach A.D., Van Berkel C., Pronk J.C., Doggett N.A., Callen D.F.;
RT "Characterization and screening for mutations of the growth arrest-
RT specific 11 (GAS11) and C16orf3 genes at 16q24.3 in breast cancer."
RL Genomics 52:325-331(1998).
RN [2]
RP MICROTUBULE-BINDING.
RX MEDLINE=20564307; PubMed=10969087;
RA Hill K.L., Hutchings N.R., Grandgenett P.M., Donelson J.E.;
RT "T lymphocyte-triggering factor of African trypanosomes is associated
RT with the flagellar fraction of the cytoskeleton and represents a new
RT family of proteins that are present in several divergent
RT eukaryotes."
RL J. Biol. Chem. 275:39369-39378(2000).
RN [3]
RP DISEASE.
RX MEDLINE=21839108; PubMed=11751847;
RA Yeh S.-D., Chen Y.-J., Chang A.C.Y., Ray R., She B.-R., Lee W.-S.,
RA Chiang H.-S., Cohen S.N., Lin-Chao S.;
RT "Isolation and properties of Gas8, a growth arrest-specific gene
RT regulated during male gametogenesis to produce a protein associated
RT with the sperm motility apparatus."
RL J. Biol. Chem. 277:6311-6317(2002).
CC -!- FUNCTION: Cytoskeletal linker that probably plays a role in the
CC spermatozoa motility (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; associates with microtubules.
CC In spermatozoa, it localizes to the flagellum (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed in the heart, skeletal muscle,
CC pancreas and liver. Weakly or not expressed in brain, placenta,
CC lung and kidney.
CC -!- DISEASE: Defects in GAS8 may be a cause of infertility in male
CC lacking gametocytes.
CC -!- SIMILARITY: Belongs to the GAS8 family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; AF050079; AAC69518.1; -
DR EMBL; AF050078; AAC69518.1; JOINED.
DR EMBL; AF050068; AAC69518.1; JOINED.
DR EMBL; AF050065; AAC69518.1; JOINED.
DR EMBL; AF050070; AAC69518.1; JOINED.
DR EMBL; AF050071; AAC69518.1; JOINED.
DR EMBL; AF050072; AAC69518.1; JOINED.
DR EMBL; AF050073; AAC69518.1; JOINED.
DR EMBL; AF050074; AAC69518.1; JOINED.
```



```

RESULT 69
CNB4_MOUSE
ID_CNB4_MOUSE STANDARD; PRT; 496 AA.
AC Q8VEG4;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Protein C14orf114 homolog.
GN C14ORF114.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Klausner R.D., Collins F.S., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Wang J., Hong L.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,
RA Diatchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Locuallano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Kulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywicki M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; BC018508; AAH18508.1; --
DR MGD; MGI:2145119; C85658.
DR InterPro; IPR002562; 3_5_exonuclease.
DR Pfam; PF01612; 3_5_exonuclease; 1.
SQ SEQUENCE 496 AA; 56785 MW; 8349C98C98002E23 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 496;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 366 HLKQQLA 372
Db 347 HLKQQLA 353
RESULT 70
G6PD_TREPA
ID_G6PD_TREPA STANDARD; PRT; 515 AA.
AC O83491;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 13-FEB-2003 (Rel. 41, Last annotation update)
DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD).
GN ZWF OR TP0478.
OS Treponema pallidum.

```

```

OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Nichols;
RA MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren B., Hardham J.M., McLeod W.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts X., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RA "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-
CC 1,5-lactone 6-phosphate + NADPH.
CC -1- PATHWAY: Pentose phosphate + NADPH.
CC -1- SIMILARITY: Belongs to the glucose-6-phosphate dehydrogenase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001224; AAC65465.1; --
DR PIR; D71319; D71319.
DR HSP; P11411; 1DPG.
DR TIGR; TP0478; --
DR InterPro; IPR001282; G6PD.
DR Pfam; PF00479; G6PD; 1.
DR Pfam; PF02781; G6PD; 1.
DR PRINTS; PR00079; G6PDHGNASE.
DR PRODOM; PD001129; G6PD; 1.
DR TIGRFAMS; TIGR00871; zwf; 1.
DR PROSITE; PS00069; G6P_DEHYDROGENASE; 1.
KW Oxidoreductase; NADP; Glucose metabolism; Complete proteome.
FT ACT SITE 194 194 BY SIMILARITY.
SQ SEQUENCE 515 AA; 58032 MW; 22936A1609EBD798 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 515;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 48 QSGALRD 54
Db 241 QSGALRD 247
RESULT 71
TRAS_HUMAN
ID TRAS_HUMAN STANDARD; PRT; 557 AA.
AC O00423;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE TNF receptor associated factor 5.
GN TRAF5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.; AND INTERACTION WITH TNFRSF5; TNFRSF8 AND LTR8.
RX MEDLINE=98172745; PubMed=9511754;
RA Mizushima S.-I., Fujita M., Ishida T., Azuma S., Kato K., Hirai M.,
RA Otsuka M., Yamamoto T., Inoue J.-I.;
RA "Cloning and characterization of a cDNA encoding the human homolog of

```

- RT tumor necrosis factor receptor-associated factor 5 (TRAF5).";
RL Gene 207:135-140(1998).
RN
RC SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKean P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton B., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smalrus D.E.,
RA Schnerch A., Schein J.E., Jones S.O.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN
RP
RX SEQUENCE OF 20-557 FROM N.A.
RX MEDLINE=97321041; PubMed=9177772;
RA Nakano H., Shindo M., Yamada K., Yoshida M.C., Santee S.M., Ware C.F.,
RA Jenkins N.A., Gilbert D.J., Yagita H., Copeland N.G., Okumura K.;
RT "Human TNF receptor-associated factor 5 (TRAF5): cDNA cloning,
RT expression and assignment of the TRAF5 gene to chromosome 1q32.";
RL Genomics 42:26-32(1997).
RN
RP
RX MEDLINE=21278352; PubMed=11384837;
RA Wajant H., Henkler F., Scheurich P.;
RT "The TNF-receptor-associated factor family: scaffold molecules for
RT cytokine receptors, kinases and their regulators.";
RL Cell. Signal. 13:389-400(2001).
RN
RP
RX MEDLINE=21519158; PubMed=11607847;
RA Bradley J.R., Pober J.S.;
RT "Tumor necrosis factor receptor-associated factors (TRAFs).";
RL Oncogene 20:6482-6491(2001).
RN
RP
RX INTERACTION WITH TNFRSF14.
RX MEDLINE=97306297; PubMed=9162022;
RA Marsters S.A., Ayres T.M., Skubatch M., Gray C.L., Rothe M.,
RA Ashkenazi A.;
RT "Herpesvirus entry mediator, a member of the tumor necrosis factor
RT receptor (TNFR) family, interacts with members of the TNFR-associated
RT factor family and activates the transcription factors NF-kappaB and
RT AP-1.";
RL J. Biol. Chem. 272:14029-14032(1997).
RN
RP
RX INTERACTION WITH TNFRSF14.
RX MEDLINE=97298041; PubMed=9153189;
RA Heu H., Solovyev I., Colombero A., Elliott R., Kelley M., Boyle W.J.;
RT "ATAR, a novel tumor necrosis factor receptor family member, signals
RT through TRAF2 and TRAF5.";
RL J. Biol. Chem. 272:13471-13474(1997).
RN
RP
RX INTERACTION WITH MAP3K14.
RX MEDLINE=97420762; PubMed=9275204;
RA Song H.Y., Regnier C.H., Kirschning C.J., Goeddel D.V., Rothe M.;
RT "Tumor necrosis factor (TNF)-mediated kinase cascades: bifurcation of
RT nuclear factor-kappaB and c-jun N-terminal kinase (JNK/SAPK) pathways
RT at TNF receptor-associated factor 2.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:9792-9796(1997).
RN
RP
RX INTERACTION WITH TNFRSF13B.
RX MEDLINE=20341628; PubMed=10880535;
RA Xia X.-Z., Treanor J., Senaldi G., Khare S.D., Boone T., Kelley M.,
RA Theill L.E., Colombero A., Solovyev I., Lee F., McCabe S., Elliott R.,
RA Miner K., Hawkins N., Guo J., Stolina M., Yu G., Wang J., Delaney J.,
RA Meng S.Y., Boyle W.J., Heu H.;
RT "TAC1 is a TRAF-interacting receptor for TAIL-1, a tumor necrosis
RT factor family member involved in B cell regulation.";
RL J. Exp. Med. 192:137-143(2000).
RN
RP
RX INTERACTION WITH TNFRSF8.
RX MEDLINE=97152965; PubMed=8998988;
RA Aizawa S., Nakano H., Ishida T., Horie R., Nagai M., Ito K.,
RA Yagita H., Okumura K., Inoue J.-I., Watanabe T.;
RT "Tumor necrosis factor receptor-associated factor (TRAF) 5 and TRAF2
RT are involved in CD30-mediated NF-kappaB activation.";
RL J. Biol. Chem. 272:2042-2045(1997).
RN
RP
RX INTERACTION WITH MAP3K5.
RX MEDLINE=98448103; PubMed=9774977;
RA Nishitoh H., Saitoh K., Mochida Y., Takeda K., Nakano H., Rothe M.,
RA Miyazono K., Ichijo H.;
RT "ASK1 is essential for JNK/SAPK activation by TRAF2.";
RL Mol. Cell 2:389-395(1998).
RN
RP
RX INTERACTION WITH TNFRSF5 AND TRAF3.
RX MEDLINE=98384149; PubMed=9718306;
RA Pullen S.S., Miller H.G., Everdeen D.S., Dang T.T., Crute J.J.,
RA Kehry M.R.;
RT "CD40-tumor necrosis factor receptor-associated factor (TRAF)
RT interactions: regulation of CD40 signaling through multiple TRAF
RT binding sites and TRAF hetero-oligomerization.";
RL Biochemistry 37:11836-11845(1998).
RN
RP
RX INTERACTION WITH TNFRSF4.
RX MEDLINE=98157982; PubMed=9488716;
RA Kawamata S., Hori T., Imura A., Takaori-Kondo A., Uchiyama T.;
RT "Activation of OX40 signal transduction pathways leads to tumor
RT necrosis factor receptor-associated factor (TRAF) 2- and
RT TRAF5-mediated NF-kappaB activation.";
RL J. Biol. Chem. 273:5808-5814(1998).
RN
RP
RX INTERACTION WITH RIPK2.
RX MEDLINE=98307936; PubMed=9642260;
RA McCarthy J.V., Ni J., Dixit V.M.;
RT "RIP2 is a novel NF-kappaB-activating and cell death-inducing
RT kinase.";
RL J. Biol. Chem. 273:16968-16975(1998).
RN
RP
RX INTERACTION WITH TNFRSF19.
RX MEDLINE=20270246; PubMed=10809768;
RA Eby M.T., Jasmin A., Kumar A., Sharma K., Chaudhary P.M.;
RT "TAJ, a novel member of the tumor necrosis factor receptor family,
RT activates the c-Jun N-terminal kinase pathway and mediates
RT caspase-independent cell death.";
RL J. Biol. Chem. 275:15336-15342(2000).
RN
RP
RX INTERACTION WITH TNFRSF11A.
RX MEDLINE=98447691; PubMed=9774460;
RA Kong B.R., Josien R., Lee S.Y., Vologodskaya M., Steinman R.M.,
RA Choi Y.;
RT "The TRAF family of signal transducers mediates NF-kappaB activation
RT by the TRANCE receptor.";
RL J. Biol. Chem. 273:28355-28359(1998).
RN
RP
RX INTERACTION WITH TTRAP.
RX MEDLINE=20309820; PubMed=10764746;
RA Eype S., Declercq W., Ibrahim A., Michiels C., Van Rietachoten J.G.,
RA Dewulf N., de Boer M., Vandenberghe P., Huybrechts D., Remacle J.E.;
RT "TTRAP, a novel protein that associates with CD40, tumor necrosis
RT factor (TNF) receptor-75 and TNF receptor-associated factors (TRAFs),
RT and that inhibits nuclear factor-kappa B activation.";
RL J. Biol. Chem. 275:18586-18593(2000).
RN
RP
RX INTERACTION WITH TNFRSF13B.
RX MEDLINE=20341628; PubMed=10880535;
RA Xia X.-Z., Treanor J., Senaldi G., Khare S.D., Boone T., Kelley M.,
RA Theill L.E., Colombero A., Solovyev I., Lee F., McCabe S., Elliott R.,
RA Miner K., Hawkins N., Guo J., Stolina M., Yu G., Wang J., Delaney J.,
RA Meng S.Y., Boyle W.J., Heu H.;
RT "TAC1 is a TRAF-interacting receptor for TAIL-1, a tumor necrosis
RT factor family member involved in B cell regulation.";
RL J. Exp. Med. 192:137-143(2000).
RN

```

(18)
RN INTERACTION WITH TNFRSF17.
RX MEDLINE=20381353; PubMed=10908663;
RA Shu H.-B., Johnson H.;
RT "B cell maturation protein is a receptor for the tumor necrosis factor
RL family member TALL-1."
CC -1- FUNCTION: Adapter protein and signal transducer that links members
CC of the tumor necrosis factor receptor family to different
CC signaling pathways by association with the receptor cytoplasmic
CC domain and kinases. Mediates activation of NF-kappa-B and probably
CC domain. Seems to be involved in apoptosis.
CC -1- SUBUNIT: Homotrimer (Probable). Heteromer with TRAF3. Associates
CC with TNFRSF5/CD40 through interaction with TRAF3. Associates with
CC LTRB/TNFRSF3, TNFRSF3, TNFRSF4, TNFRSF8/CD30, TNFRSF11A/RANK,
CC TNFRSF13B/TAC1, TNFRSF14, TNFRSF17, TNFRSF19/TROY, RIPK2, MAP3K14,
CC MAP3K5, and TRAF and TNF receptor associated protein TRAP.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- TISSUE SPECIFICITY: Expressed in spleen, thymus, prostate, testis,
CC ovary, small intestine, colon, and peripheral blood.
CC -1- DOMAIN: The MATH/TRAF domain binds to receptor cytoplasmic
CC domains.
CC -1- SIMILARITY: Contains 1 MATH domain.
CC -1- SIMILARITY: Contains 1 RING-type zinc finger.
CC -1- SIMILARITY: Contains 2 TRAF-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB000509; BAA25262.1; -.
CC EMBL; BC029600; AAH29600.1; -.
CC EMBL; U69108; AAC51329.1; -.
CC
CC Query Match 1.3%; Score 7; DB 1; Length 557;
CC Best Local Similarity 100.0%; Pred. No. 1.3e+02;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 406 IKKLEKE 412
CC Db 285 IKKLEKE 291
CC
CC RESULT 72
CC IF2_THETH STANDARD; PRT; 571 AA.
CC AC P48515;
CC DT 01-FEB-1996 (Rel. 33, Created)
CC DT 01-FEB-1996 (Rel. 33, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Translation initiation factor IF-2.
CC GN INFB.
CC OS Thermus thermophilus.
CC OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
CC OC Thermus.
CC OX NCBI_TaxID=274;
CC RN [1]
CC RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
CC RC STRAIN=HB8 / ATCC 27634;
CC RX MEDLINE=97182584; PubMed=9030723;
CC RA Vornlocher H.P., Scheible W.R., Faulhammer H.G., Sprinzl M.;
CC RT "Identification and purification of translation initiation factor 2
CC (IF2) from Thermus thermophilus";
CC RL Eur. J. Biochem. 243:66-71(1997).
CC
CC -1- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).

```

```

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the IF-2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z48001; CA888038.1; -.
CC EMBL; S52276; S52276.
CC HAMAP; MF_00100; -.
CC InterPro; IPR000795; EF_GTPbind.
CC InterPro; IPR004161; EFTU_D2.
CC InterPro; IPR000178; IP2_N.
CC InterPro; IPR006847; IP2_N.
CC InterPro; IPR005325; Small_GTP.
CC InterPro; IPR009000; Translat_factor.
CC Pfam; PF03144; GTP_EFTU; 1.
CC Pfam; PF04760; IF2_N; 1.
CC Pfam; PF04760; IF2_N; 1.
CC TIGRFAMs; TIGR00487; IF-2; 1.
CC TIGRFAMs; TIGR00231; small_GTP; 1.
CC PROSITE; PS01176; IF2; 1.
CC KW Initiation factor; Protein biosynthesis; GTP-binding.
CC FT DOMAIN 74 222 G-DOMAIN.
CC FT NP_BIND 80 87 GTP (BY SIMILARITY).
CC FT NP_BIND 126 130 GTP (BY SIMILARITY).
CC FT NP_BIND 180 183 GTP (BY SIMILARITY).
CC SQ SEQUENCE 571 AA; 63190 MW; 7004210C19C90F01 CRC64;
CC
CC Query Match 1.3%; Score 7; DB 1; Length 571;
CC Best Local Similarity 100.0%; Pred. No. 1.3e+02;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 247 AREEEK 253
CC Db 345 AREEEK 351
CC
CC RESULT 73
CC VIPC_SALTI STANDARD; PRT; 578 AA.
CC AC Q04975;
CC DT 01-NOV-1995 (Rel. 32, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Vi polysaccharide biosynthesis protein vipC/tviE.
CC GN VIPC OR TVIE OR STY4656 OR T4349.
CC OS Salmonella typhi.
CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC OC Enterobacteriaceae; Salmonella.
CC OX NCBI_TaxID=601;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=Ty2 / ATCC 700931;
CC RX MEDLINE=94069051; PubMed=8248629;
CC RA Waxlin H., Virlogeux I., Kolyva S., Popoff M.Y.;
CC RT "Identification of six open reading frames in the Salmonella enterica
CC subsp. enterica ser. Typhi viab locus involved in Vi antigen
CC production.";
CC RL Res. Microbiol. 144:363-371(1993).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=GIU 10007;
CC RX MEDLINE=93322324; PubMed=8331073;
CC RA Hashimoto Y., Li N., Yokoyama H., Ezaki T.;
CC RT "Complete nucleotide sequence and molecular characterization of Viab
CC region encoding Vi antigen in Salmonella typhi.";
CC RL J. Bacteriol. 175:4456-4465(1993).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krog A., Larsen T.S., Leather S., Moulé S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whithead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=TY2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liu S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyan V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
RN -1- PATHWAY: Vi polysaccharide biosynthesis.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67785; CAA47994.1; -
DR EMBL; D14156; BAA03195.1; -
DR EMBL; AL627283; CAA06776.1; -
DR EMBL; AB016848; AB071802.1; -
DR PIR; E36892; E36892.
DR PIR; S28492; S28492.
DR InterPro; IPR001296; Glyco_transf_1.
DR Pfam; PF00534; Glycos_transf_1.
KW Complete proteome.
FT CONFLICT 290 290 S -> G (IN REF. 1 AND 4).
FT CONFLICT 362 363 KH -> ND (IN REF. 1).
FT CONFLICT 362 363
SQ SEQUENCE 578 AA; 65009 MW; 8D420563D868C189 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 578;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 EEKLAAL 176
DB 5 EEKLAAL 11

RESULT 74
Y085 CHLTR
ID Y085 CHLTR STANDARD; PRT; 579 AA.
AC G84087;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CT085.
GN CT085.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=9900809; PubMed=97841136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
```

```

RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- SIMILARITY: Belongs to the ubiD family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001283; AAC67676.1; -
DR PIR; E71557; E71557.
DR InterPro; IPR002830; carboxylase.
DR Pfam; PF01977; UbiD; 1.
DR TIGRfams; TIGR00148; TIGR00148; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 579 AA; 65354 MW; E2512FC314216315 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 579;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 486 ERRPEGP 492
DB 268 ERRPEGP 274

RESULT 75
Y454 CHLCV
ID Y454 CHLCV STANDARD; PRT; 581 AA.
AC Q34023;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 42, Last annotation update)
DE Hypothetical protein CCA00454 (ORF).
GN CCA00454.
OS Chlamydia caviae.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83557;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=97426043; PubMed=9282747;
RA Hsia R.C., Pannekoek Y., Ingerowski E., Bavoil P.M.;
RT "Type III secretion genes identify a putative virulence locus of
RT Chlamydia."
RL Mol. Microbiol. 25:351-359(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=22569155; PubMed=12682364;
RA Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA Heidelberg J., Holtzapple E., Khouri H., Federova N.B., Carty H.A.,
RA Umayam L.A., Haft D.H., Peterson J., Bean M.J., White O.,
RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoil P.M.,
RA Frazer C.M.;
RT "Genome sequence of Chlamydia caviae (Chlamydia psittaci) GPIC":
RT examining the role of niche-specific genes in the evolution of the
RT Chlamydiaceae."
RL Nucleic Acids Res. 31:2134-2147(2003).
RN [3]
RP SEQUENCE FROM N.A.
CC -1- SIMILARITY: Belongs to the ubiD family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
```

```

CC or send an email to license@sib.ch).
CC -----
DR EMBL; U88070; AAB71513.1; -.
DR EMBL; AE016995; AAP05200.1; -.
DR TIGR; CCA00454; -.
DR InterPro; IPR002830; carboxylase.
DR Pfam; PF01977; UniD; 1.
DR TIGRFAMs; TIGR00148; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 581 AA; 66207 MW; 8D51F340DD99A2DC CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 581;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 486 ERRPEGP 492
Db 268 ERRPEGP 274

RESULT 76
D170 HUMAN
ID D170 HUMAN STANDARD; PRT; 585 AA.
AC Q16204;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE H4 protein.
GN D108170.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=94336206; PubMed=8058316;
RA Grice M., Carrato A., Santoro M., Fusco A., Melillo R.M., Vecchio G.;
RT "Cloning and characterization of H4 (D108170), a gene involved in RET
rearrangements in vivo.";
RL Oncogene 9:2531-2535(1994).
CC -!- SURCELLULAR LOCATION: May be a cytoskeletal protein.
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed.
CC -!- DOMAIN: The protein has mostly an alpha helical conformation
similar to myosin heavy-chain tail that might adopt a coiled-coil
conformation.
CC -!- DISEASE: ABOUT 20% OF HUMAN THYROID PAPILLARY CARCINOMAS (PACT)
SEEM TO BE CAUSED BY A REARRANGEMENT INV(10)(Q11.2Q21) INVOLVING
RET AND H4(D108170) THAT GENERATES THE RET/PTC1 ONCOGENE.
CC -!- SIMILARITY: TO C.ELEGANS T0989.4.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib.ch).

EMBL; S72869; AAC0637.1; -.
DR PIR; I58403; I58403.
DR MIM; 601985; -.
DR MIM; 188550; -.
GO: GO:0005200; F:structural constituent of cytoskeleton; TAS.
KW Proto-oncogene; Chromosomal translocation; SH3-binding; Repeat.
FT DOMAIN 17 20 POLY-SER.
FT DOMAIN 32 44 POLY-GLY.
FT DOMAIN 440 448 POLY-PRO.
FT DOMAIN 106 235 5 X 29 AA TANDEM REPEATS.
FT REPEAT 106 134 1.
FT REPEAT 135 163 2.
FT REPEAT 164 192 3.
FT REPEAT 193 206 4 (APPROXIMATE).

```

```

FT REPEAT 207 235 5.
FT SITE 442 451 SH3-BINDING (POTENTIAL).
SQ SEQUENCE 585 AA; 65916 MW; 33FD9C81DA72AA8A CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 585;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KXIKLE 410
Db 176 KXIKLE 182

RESULT 77
YM72 YEAST
ID YM72 YEAST STANDARD; PRT; 590 AA.
AC Q05021;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 67.6 kDa protein in MRPL44-MTF1 intergenic region.
GN YMR227C OR YMR959.09C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97333268; PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
XIII.";
RL Nature 387:90-93(1997).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).

EMBL; Z49933; CAA90198.1; -.
DR PIR; S57594; S57594.
DR Germonline; 142902; -.
DR TRANSFAC; T03088; -.
DR SGD; S0004840; TAF67.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0000114; P:G1-specific transcription in mitotic cell c.; IPI.
DR InterPro; IPR006751; TAF155_N.
DR Pfam; PF04658; TAF155_N; 1.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 50 53 POLY-LYS.
FT DOMAIN 200 203 POLY-GLU.
FT DOMAIN 368 373 POLY-GLU.
FT DOMAIN 413 421 POLY-ASP.
FT DOMAIN 517 531 POLY-GLU.
FT DOMAIN 427 549 COILED COIL (POTENTIAL).
SQ SEQUENCE 590 AA; 67555 MW; C014E7419B0B1C61 CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 590;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 IKKLEKE 412
Db 479 IKKLEKE 485

```



```

RESULT 78
DNJM_MYCGE
ID DNJM_MYCGE STANDARD; PRT; 601 AA.
AC P4742; Q49288;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE DnaJ-like protein MG200.
GN MG200.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uettersback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Boff K.P., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
RN [2]
SEQUENCE OF 281-409 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Boff K.P., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing."
RL J. Bacteriol. 175:7918-7930(1993).
CC -!- SIMILARITY: Contains 1 J domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U39699; AAC71418.1; -
CC DR EMBL; U02163; AAD12445.1; -
CC DR PIR; A64222; A64222.
CC DR HSP; P08622; LBQZ.
CC TIGR; MG200; -
CC DR InterPro; IPR001623; DnaJ_N.
CC DR InterPro; IPR003095; Hsp_DnaJ.
CC DR Pfam; PF00226; DnaJ_1.
CC DR PRINTS; PR00625; DnaJPROTEIN.
CC DR SMART; SM00271; DnaJ_1.
CC DR PROSITE; PS00636; DnaJ_1; 1.
CC DR PROSITE; PS00076; DnaJ_2; 1.
CC KW Hypothetical protein; Chaperone; Complete proteome.
CC FT DOMAIN
CC SEQUENCE 601 AA; 68537 MW; P9FAE352E3341D093 CRC64;
Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 601;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 106 EPEPTPV 112
Db 215 EPEPTPV 221
|||||
|||||
RESULT 79
SP2_HUMAN
ID SP2_HUMAN STANDARD; PRT; 606 AA.
AC Q02086;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor Sp2.
GN SP2 OR KIAA0048.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Suzuki T., Tanaka A., Sato S.,
RA Seki N., Kawatabayashi Y., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1."
RNA Res. 1:223-229(1994).
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=Lung, Lymph, and Spleen;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold Z.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Mang J., Hsieh P.,
RA Diatchenko M., Soares M.B., Bonaldo M.P., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krawinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
SEQUENCE OF 112-606 FROM N.A.
RX MEDLINE=93024366; PubMed=1341900;
RA Kingsley C., Winoto A.;
RT "Cloning of GT box-binding proteins: a novel Sp1 multigene family
RT regulating T-cell receptor gene expression."
RL Mol. Cell. Biol. 12:4251-4261(1992).
CC -!- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY
CC ACTIVATES MENA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL
CC RECOGNITION SITES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D28588; BAA05923.1; -
CC DR EMBL; BC016680; AAH16680.1; -
CC DR EMBL; BC033814; AAH33814.1; -
CC DR EMBL; M97190; AAA36629.1; -
CC DR PIR; A44489; A44489.
CC DR HSP; P08047; 1SP2.
CC DR TRANSFAC; T02356; -
CC DR Genew; HGNC:11207; SP2.
CC MIN; 601801; -

```


DR GO: GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.
DR GO: GO:0006955; P:immune response; TAS.
DR GO: GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00096; ZF-C2H2; 3.
DR ProDom: PD000003; Znf_C2H2; 2.
DR SMART: SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT ZN_FING 518 542 C2H2-TYPE 1.
FT ZN_FING 548 572 C2H2-TYPE 2.
FT ZN_FING 578 600 C2H2-TYPE 3.
SQ SEQUENCE 606 AA; 64153 MW; A27C6D460D36E186 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 606;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 IQLQMEQ 274
DB 475 IQLQMEQ 481

RESULT 80
ID DNAK_BUCBP STANDARD; PRT; 638 AA.
AC Q8K9Y8;
DR 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).
GN DNAK OR BUSG146.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Wernegreen J.J., Sandstrom J.P., Moran N.A., Anderson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria."; Science 296:2376-2379(2002).
CC -1- FUNCTION: Acts as a chaperone (By similarity).
CC -1- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC -1- SIMILARITY: Belongs to the heat shock protein 70 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; AE014090; AAM67714.1; .
CC InterPro: IPR001023; Hsp70.
CC Pfam; PF00012; HSP70; 1.
CC PRINTS; PR00301; HEATSHOCK70.
CC ProDom; PD000089; HSP70; 1.
CC PROSITE; PS00297; HSP70_1; 1.
CC PROSITE; PS00329; HSP70_2; 1.
CC PROSITE; PS01036; HSP70_3; 1.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation;
KW Complete proteome.
FT MOD_RES 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 638 AA; 70344 MW; 541CCFE2581F043F CRC64;

Query Match 1.3%; Score 7; DB 1; Length 638;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 RSKLES 228
DB 302 RSKLES 308

RESULT 81
ID DNAK_BUCBP STANDARD; PRT; 638 AA.
AC P59565;
DR 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).
GN DNAK OR BBP142.
OS Buchnera aphidicola (subsp. Baizongia pistaciae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=135842;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426901; PubMed=12522265;
RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,
RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
RT "Reductive genome evolution in Buchnera aphidicola."; Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
CC -1- FUNCTION: Acts as a chaperone (By similarity).
CC -1- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC -1- SIMILARITY: Belongs to the heat shock protein 70 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; AE014016; AAO26876.1; .
CC HAMAP; MF 00332; .
CC InterPro: IPR001023; Hsp70.
CC Pfam; PF00012; HSP70; 1.
CC PRINTS; PR00301; HEATSHOCK70.
CC ProDom; PD000089; HSP70; 1.
CC PROSITE; PS00297; HSP70_1; 1.
CC PROSITE; PS00329; HSP70_2; 1.
CC PROSITE; PS01036; HSP70_3; 1.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation;
KW Complete proteome.
FT MOD_RES 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 638 AA; 70368 MW; 9C4F87571EE0AD82 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 638;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 RSKLES 228
DB 302 RSKLES 308

RESULT 82
ID DNAK_PSESG STANDARD; PRT; 638 AA.
AC Q9WMG9;
DR 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).
GN DNAK OR BUSG146.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Wernegreen J.J., Sandstrom J.P., Moran N.A., Anderson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria."; Science 296:2376-2379(2002).
CC -1- FUNCTION: Acts as a chaperone (By similarity).
CC -1- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC -1- SIMILARITY: Belongs to the heat shock protein 70 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; AE014090; AAM67714.1; .
CC InterPro: IPR001023; Hsp70.
CC Pfam; PF00012; HSP70; 1.
CC PRINTS; PR00301; HEATSHOCK70.
CC ProDom; PD000089; HSP70; 1.
CC PROSITE; PS00297; HSP70_1; 1.
CC PROSITE; PS00329; HSP70_2; 1.
CC PROSITE; PS01036; HSP70_3; 1.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation;
KW Complete proteome.
FT MOD_RES 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 638 AA; 70344 MW; 541CCFE2581F043F CRC64;

Query Match 1.3%; Score 7; DB 1; Length 638;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

DE protein) (HSP70).
GN DNAX.
OS Pseudomonas syringae (pv. glycinea).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
CX NCBI_TaxID=318;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PG4180;
RA MEDLINE=99407915; PubMed=10478477;
RX Keith L.M.W., Partridge J.E., Bender C.L.;
RT "dnak and the heat stress response of Pseudomonas syringae pv.
RL glycinea.";
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF135163; AAD31868.1; -.
CC HSP; P04475; IDKX.
CC HAMAP; MF_00332; -. 1.
CC InterPro; IPR001023; Hsp70.
CC Pfam; PF00012; HSP70; 1.
CC PRINTS; PR00301; HEATSHOCK70.
CC ProDom; PD000089; Hsp70; 1.
CC PROSITE; PS00297; HSP70_1; 1.
CC PROSITE; PS00329; HSP70_2; 1.
CC PROSITE; PS01036; HSP70_3; 1.
CC Chapterone; ATP-binding; Heat shock; Phosphorylation.
FT MOD RES 199 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT SEQUENCE 638 AA; 68897 MW; DE2C34D287FC21B CRC64;
SQ
Query Match 1.3%; Score 7; DB 1; Length 638;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 222 RSKLES 228
DB 302 RSKLES 308
|||||
RESULT 83
ID DNAX_PSESM STANDARD; PRT; 638 AA.
AC Q87NF0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Chapterone protein dnak (Heat shock protein 70) (Heat shock 70 kDa
DE protein) (HSP70).
GN DNAX OR PSPT04505.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
CX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=22834015; PubMed=12928499;
RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
RA Winn M.L., Dodson R.J., Deboy R.T., Durkin A.S., Kolonay J.F.,
RA Madupu R., Daugherty S., Brinkac L., Beanan M.J., Haft D.H.,
RA Nelson W.C., Davidson T., Zafar N., Zhou L., Liu J., Yuan Q.,
RA Khouri H., Fedorova N., Tran B., Russell D., Berry K., Utterback T.,
RA Van Aken S.E., Feldblyum T.V., D'Ascenzo M., Deng W.-L., Ramos A.R.,
protein) (HSP70).
GN DNAX.
OS Pseudomonas syringae (pv. glycinea).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
CX NCBI_TaxID=318;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PG4180;
RA MEDLINE=99407915; PubMed=10478477;
RX Keith L.M.W., Partridge J.E., Bender C.L.;
RT "dnak and the heat stress response of Pseudomonas syringae pv.
RL glycinea.";
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF016872; AA057953.1; -.
CC TIGR; PSP04505; -.
CC HAMAP; MF_00332; -. 1.
CC InterPro; IPR001023; Hsp70.
CC Pfam; PF00012; HSP70; 1.
CC PRINTS; PR00301; HEATSHOCK70.
CC ProDom; PD000089; Hsp70; 1.
CC PROSITE; PS00297; HSP70_1; 1.
CC PROSITE; PS00329; HSP70_2; 1.
CC PROSITE; PS01036; HSP70_3; 1.
CC Chapterone; ATP-binding; Heat shock; Phosphorylation;
KW Complete proteome.
FT MOD RES 199 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT SEQUENCE 638 AA; 68755 MW; E6504919AA678BCC CRC64;
SQ
Query Match 1.3%; Score 7; DB 1; Length 638;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 222 RSKLES 228
DB 302 RSKLES 308
|||||
RESULT 84
ID UVRB_CLOPE STANDARD; PRT; 659 AA.
AC Q46323;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE UvrABC system protein B (UvrB protein) (Excinuclease ABC subunit B).
GN UVRB OR CPE0303.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
CX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [2]
RP SEQUENCE OF 69-155 FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=96033407; PubMed=7559358;
RA Katayama S.-I., Dupuy B., Garnier T., Cole S.T.;
RT "Rapid expansion of the physical and genetic map of the chromosome of
RT Clostridium perfringens CPN50.";
RJ Bacteriol. 177:5680-5685(1995).
CC -!- FUNCTION: The UvrABC repair system catalyzes the recognition and
CC processing of DNA lesions. A damage recognition complex composed

```

```

CC of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. Upon
CC binding of the uvrA(2)B(2) complex to a putative damaged site, the
CC DNA wraps around one uvrB monomer. DNA wrap is dependent on ATP
CC binding by uvrB and probably causes local melting of the DNA
CC helix, facilitating insertion of uvrB beta-hairpin between the DNA
CC strands. Then uvrB probes one DNA strand for the presence of a
CC lesion. If a lesion is found the uvrA subunits dissociate and the
CC uvrB-DNA preincision complex is formed. This complex is
CC subsequently bound by uvrC and the second uvrB is released. If no
CC lesion is found, the DNA wraps around the other uvrB subunit that
CC will check the other stand for damage (By similarity).
CC -!- SUBUNIT: Forms a heterotetramer with uvrA during the search for
CC lesions. Interacts with uvrC in an incision complex (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- DOMAIN: The beta-hairpin motif is involved in DNA binding (By
CC similarity).
CC -!- SIMILARITY: Contains 1 UVR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC -----
CC EMBL; AP003186; BAB80009.1; -.
CC EMBL; X86531; CAAG0243.1; -.
CC HSSP; P56981; ID9X.
CC HAMAP; MF_00204; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR004807; UvrB.
CC InterPro; IPR001943; UvrB/C.
CC Pfam; PF00271; Helicase_C; 1.
CC Pfam; PF02151; UVR; 1.
CC SMART; SM00487; DEXDC; 1.
CC SMART; SM00490; HELICC; 1.
CC TIGRFAMs; TIGR00631; uvrB; 1.
CC PROSITE; PS0151; UVR; 1.
CC SOS response; Excision nuclease; DNA repair; DNA recombination;
KW DNA excision; ATP-binding; Complete proteome.
FT NP_BIND 38 45 ATP (POTENTIAL).
FT DOMAIN 91 114 BETA-HAIRPIN.
FT DOMAIN 622 657 UVR.
FT CONFLICT 69 75 AAQLCAE -> GSALIVK (IN REF. 2).
SQ SEQUENCE 659 AA; 75523 MW; 4F0D94B3C647B01A CRC64;

Query Match 1.3%; Score 7; DB 1; Length 659;
Best Local Similarity 100.0%; Pred.No.1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 KLIQVYE 301
DB 623 KLIQVYE 629

RESULT 85
UVRB METAC
ID UVRB METAC STANDARD; PRT; 670 AA.
AC Q8TK53;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE UVRABC system protein B (UvrB protein) [Excision nuclease ABC subunit B].
GN UVRB OR MA3323.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_taxid=2214;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=1193238;
RA Galagan J.E., Nushbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA FitzRugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf M.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
CC -!- FUNCTION: The UVRABC repair system catalyzes the recognition and
CC processing of DNA lesions. A damage recognition complex composed
CC of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. Upon
CC binding of the uvrA(2)B(2) complex to a putative damaged site, the
CC DNA wraps around one uvrB monomer. DNA wrap is dependent on ATP
CC binding by uvrB and probably causes local melting of the DNA
CC helix, facilitating insertion of uvrB beta-hairpin between the DNA
CC strands. Then uvrB probes one DNA strand for the presence of a
CC lesion. If a lesion is found the uvrA subunits dissociate and the
CC uvrB-DNA preincision complex is formed. This complex is
CC subsequently bound by uvrC and the second uvrB is released. If no
CC lesion is found, the DNA wraps around the other uvrB subunit that
CC will check the other stand for damage (By similarity).
CC -!- SUBUNIT: Forms a heterotetramer with uvrA during the search for
CC lesions. Interacts with uvrC in an incision complex (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- DOMAIN: The beta-hairpin motif is involved in DNA binding (By
CC similarity).
CC -!- SIMILARITY: Contains 1 UVR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC -----
CC EMBL; AE011036; AAM06692.1; -.
CC HAMAP; MF_00204; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR004807; UvrB.
CC InterPro; IPR001943; UvrB/C.
CC Pfam; PF00271; Helicase_C; 1.
CC Pfam; PF02151; UVR; 1.
CC SMART; SM00487; DEXDC; 1.
CC SMART; SM00490; HELICC; 1.
CC TIGRFAMs; TIGR00631; uvrB; 1.
CC PROSITE; PS0151; UVR; 1.
CC SOS response; Excision nuclease; DNA repair; DNA recombination;
KW DNA excision; ATP-binding; Complete proteome.
FT NP_BIND 64 71 ATP (POTENTIAL).
FT DOMAIN 117 140 BETA-HAIRPIN.
FT DOMAIN 631 665 UVR.
SQ SEQUENCE 670 AA; 77788 MW; 4BB8ED22CB97188B CRC64;

Query Match 1.3%; Score 7; DB 1; Length 670;
Best Local Similarity 100.0%; Pred.No.1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 IKKLEKE 412
DB 660 IKKLEKE 666

```

```

RESULT 86
ID UVRB METWA STANDARD; PRT; 670 AA.
AC QSPRZ9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE UvrABC system protein B (UvrB protein) (Excinuclease ABC subunit B).
GN UVRB OR MM3289.
OS Methanosarcina mazelii (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Goel / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=1212524;
RA Deppenmeier U., Johann A., Hartesch T., Markl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wierze A., Baeumer S., Jacobi C.,
RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazelii: evidence for lateral gene
transfer between Bacteria and Archaea.";
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC -!- FUNCTION: The UvrABC repair system catalyzes the recognition and
processing of DNA lesions. A damage recognition complex composed
of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. Upon
binding of the uvrA(2)B(2) complex to a putative damaged site, the
DNA wraps around one uvrB monomer. DNA wrap is dependent on ATP
binding by uvrB and probably causes local melting of the DNA
helix, facilitating insertion of uvrB beta-hairpin between the DNA
strands. Then uvrB probes one DNA strand for the presence of a
lesion. If a lesion is found the uvrA subunits dissociate and the
uvrB-DNA preincision complex is formed. This complex is
subsequently bound by uvrC and the second uvrB is released. If no
lesion is found, the DNA wraps around the other uvrB subunit that
will check the other strand for damage (By similarity).
CC -!- SUBUNIT: Forms a heterotetramer with uvrA during the search for
similarity). Interacts with uvrC in an incision complex (By
similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- DOMAIN: The beta-hairpin motif is involved in DNA binding (By
similarity).
CC -!- SIMILARITY: Belongs to the uvrB family.
CC -!- SIMILARITY: Contains 1 UVR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AB013589; AM32985.1; -.
CC HAMAP; MF_00204; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR004807; UvrB.
CC InterPro; IPR001943; UvrB/C.
CC Pfam; PF00271; helicase_C; 1.
CC Pfam; PF02151; UVR; 1.
CC TIGRFAMs; TIGR00631; uvrB; 1.
CC PROSITE; P55015; UVR; 1.
CC SOS response; Excision nuclease; DNA repair; DNA recombination;
KW DNA excision; ATP-binding; Complete proteome.
FT NP BIND 64 71
FT DOMAIN 117 140 BETA-HAIRPIN.
FT DOMAIN 531 566
SQ SEQUENCE 670 AA; 77690 MW; C53D15B2210746EA CRC64;

Query Match 1.3%; Score 7; DB 1; Length 670;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 DGAQGE 88

RESULT 87
ID CA39 CHICK STANDARD; PRT; 675 AA.
AC P32017;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 3(IX) chain precursor.
GN COL9A3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=92250566; PubMed=1577778;
RA Hay-El R., Sharma Y.D., Aguilera A., Ueyama N., Wu J.J.,
RA Eyre D.R., Juricic L., Chandrasekaran S., Li M., Nah H.D.,
RA Upholt W.B., Tanzer M.L.;
RT "Cloning and developmental expression of the alpha 3 chain of chicken
type IX collagen.";
RL J. Biol. Chem. 267:10070-10076(1992).
CC -!- FUNCTION: COLLAGEN TYPE IX IS A MINOR CARTILAGE NONFIBRILLAR
COLLAGEN. IT IS ASSOCIATED WITH TYPE II COLLAGEN FIBRILS.
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(IX),
ALPHA 2(IX), AND ALPHA 3(IX).
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; M83179; AB59960.1; -.
CC InterPro; IPR008161; C1g_helix.
CC InterPro; IPR008160; Collagen.
CC Pfam; PF01391; Collagen; 11.
CC ProDom; PD000007; C1g_helix; 3.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 21
FT CHAIN 22 675 COLLAGEN ALPHA 3(IX) CHAIN.
FT DOMAIN 25 515 NONHELIICAL REGION 3 (COL3).
FT DOMAIN 516 546 NONHELIICAL REGION 3 (NC3).
FT DOMAIN 547 626 TRIPLE-HELICAL REGION 2 (COL2).
FT DOMAIN 627 631 NONHELIICAL REGION 2 (NC2).
FT DOMAIN 632 658 TRIPLE-HELICAL REGION 1 (COL1).
FT DOMAIN 659 675 NONHELIICAL REGION 1 (NC1).
FT SITE 242 244 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 591 593 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 479 479 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 675 AA; 63013 MW; C593FEC924A10098 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 675;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 DGAQGE 88
```

```

Db 631 DGAQEP 637
|||||
ID HS7E DROME STANDARD; PRT; 687 AA.
AC P29845;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Heat shock 70 kDa protein cognate 5.
GN HSC70 OR HSC70-5.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93292982; PubMed=8514184;
RA Rubin D.M., Mehta A., Zhu J., Shoham S., Chen X.J., Wells Q.,
RA Palter K.B.;
RT "Genomic structure and sequence analysis of Drosophila melanogaster
RT HSC70 genes.";
RL Gene 128:153-163(1993).
CC -!- DEVELOPMENTAL STAGE: Heat shock cognate proteins are expressed
CC constitutively during normal development.
CC -!- SIMILARITY: Belongs to the heat shock protein 70 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; L01502; AAA28628.1; -
DR PIR; JN0667; JN0667.
DR HSSP; P04475; 1DQ4.
DR FLYBase; FBgn0001220; Hsc70-5.
DR GO; GO:0003773; P:heat shock protein activity; NAS.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PRODom; PD000089; Hsp70; 1.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DX ATP-binding; Heat shock; Multigene family.
SQ SEQUENCE 687 AA; 74208 MW; 43F6BCA1D2052EFB CRC64;

Query Match 1.3%; Score 7; DB 1; Length 687;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 222 RSKLES 228
|||||
Db 348 RSKLES 354
|||||

RESULT 89
SYM_METAC
ID SYM_METAC STANDARD; PRT; 712 AA.
AC Q8TUS;
DE 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase)
DE (MetrS).
GN METG OR MA4046.
OS Methanosarcina acetivorans.

Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=1192238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Anoor D., Brown A.,
RA Alton N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Parry J.G., Jarell K.P., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson K.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
CC -!- FUNCTION: Is required not only for elongation of protein synthesis
CC but also for the initiation of all mRNA translation through
CC initiator tRNA (Met) aminoacylation (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +
CC diphosphate + L-methionyl-tRNA(Met).
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC MetG subfamily 1.
CC -!- SIMILARITY: Contains 1 tRNA-binding domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE011117; AA07394.1; -
DR HAMAP; MF 00098; fused; 1.
DR InterPro; IPR004455; MetG_Cterm.
DR InterPro; IPR008224; MetRS_dimerising.
DR InterPro; IPR008994; Nucleic acid OB.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002304; tRNA-synt_met.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR Pfam; PF01588; tRNA_bind; 1.
DR PIRSF; PIRSF001528; MetRS_dimerising; 1.
DR PRINTS; PR01041; TRNASYNTHET.
DR TIGRFAMs; TIGR00398; metG; 1.
DR TIGRFAMs; TIGR00399; metG_C term; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_1; FALSE_NEG.
DR PROSITE; PS00886; TRD; 1.
DX Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
DX Amino-binding; tRNA-binding; Metal-binding; Zinc; Complete proteome.
FT SITE 20 30 "HIGH" REGION.
FT SITE 334 338 "RMSKS" REGION.
FT DOMAIN 610 712 tRNA-BINDING.
FT METAL 151 151 ZINC (BY SIMILARITY).
FT METAL 154 154 ZINC (BY SIMILARITY).
FT METAL 163 163 ZINC (BY SIMILARITY).
FT METAL 167 167 ZINC (BY SIMILARITY).
FT BINDING 337 337 ATP (BY SIMILARITY).
SQ SEQUENCE 712 AA; 79747 MW; ECA52862B91BD25B CRC64;

Query Match 1.3%; Score 7; DB 1; Length 712;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 115 GEKEPSK 121
DB 569 GEKEPSK 575

RESULT 90
SYM METMA STANDARD; PRT; 715 AA.
AC Q9Y222; Q8WZ57; Q96PB6; Q9BS06;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase)
DE (MethRS).
DE METG OR MM0867.
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goel / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartoch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wierst A., Baumer S., Jacobi C.,
RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Frittz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
RT transfer between Bacteria and Archaea";
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC -!- FUNCTION: Is required not only for elongation of protein synthesis
CC but also for the initiation of all mRNA translation through
CC initiator tRNA(fMet) aminoacylation (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +
CC diphosphate + L-methionyl-tRNA(Met).
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC MetG subfamily 1.
CC -!- SIMILARITY: Contains 1 tRNA-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE013312; AM30563.1; ALT_INIT.
DR HAMAP; MF_00098; fused; 1.
DR InterPro; IPR004495; MetG Cterm.
DR InterPro; IPR008224; MetRS dimerising.
DR InterPro; IPR008994; Nucleic acid OB.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002104; tRNA-synt_met.
DR InterPro; IPR002547; tRNA_bind.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR Pfam; PF01588; tRNA_bind; 1.
DR Pfam; PF01589; tRNA_bind; 1.
DR PRINTS; PRS001528; MetRS dimerising; 1.
DR PRINTS; PR01041; TRNASYNTMET.
DR TIGRFAMs; TIGR00398; metG; 1.
DR TIGRFAMs; TIGR00399; metG_Cterm; 1.
DR PROSITE; PS00178; AA TRNA_LIGASE_I; FALSE_NEG.
DR PROSITE; PS00886; TRBD; 1.
KW Aminoacyl-tRNA synthetase; Metal-binding; Zinc; Complete proteome.
FT SITE 20 30 "HIGH" REGION
FT SITE 334 338 "WMSKS" REGION.

```

```

FT DOMAIN 613 715 TRNA-BINDING.
FT METAL 151 151 ZINC (BY SIMILARITY).
FT METAL 154 154 ZINC (BY SIMILARITY).
FT METAL 163 163 ZINC (BY SIMILARITY).
FT METAL 167 167 ZINC (BY SIMILARITY).
FT BINDING 337 337 ATP (BY SIMILARITY).
SQ SEQUENCE 715 AA; 79964 MW; 926A7B57855DA9CD CRC64;

Query Match 1-3%; Score 7; DB 1; Length 715;
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY 115 GEKEPSK 121
DB 569 GEKEPSK 575

RESULT 91
MTOL_HUMAN STANDARD; PRT; 717 AA.
ID MTOL_HUMAN STANDARD; Q9Y222; Q8WZ57; Q96PB6; Q9BS06;
AC Q9Y222; Q8WZ57; Q96PB6; Q9BS06;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE MTOL protein homolog (CGI-02).
DE MTOL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 5).
RX MEDLINE=2072150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Chang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of novel human genes evolutionarily conserved in
RT Caenorhabditis elegans by comparative proteomics.";
RL Genome Res. 10:703-713(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 5).
RA Li R.H., Li X.M., Zhang X., Guan W.X.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 5 AND 6).
RC TISSUE=Lymph, and Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udgin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madao A., Rodrigues S., Sanchez A.,
RA Whiting M., Madao A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Scherch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- ALTERNATIVE PRODUCTS.
CC Event=Alternative splicing; Named isoforms=5;
CC Name=3; Synonyms=4;
CC IsoId=Q9Y222-1; Sequence=Displayed;
CC Name=1;
CC IsoId=Q9Y222-2; Sequence=VSP_001749, VSP_001751;
CC Name=2;
CC IsoId=Q9Y222-3; Sequence=VSP_001750;

```



```
CC Name=5;
CC IsoId=Q9Y222-4; Sequence=VSP_001751;
CC Name=6;
CC IsoId=Q9Y222-5; Sequence=VSP_001748;
CC -!- SIMILARITY: Belongs to the gidA family.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 9.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF132937; AAD27712.1; ALT_FRAME.
CC EMBL; AF469110; AAL82394.1; -.
CC EMBL; AF469111; AAL82395.1; -.
CC EMBL; AF442963; AAL35894.1; -.
CC EMBL; AF319422; AAG2814.3; -.
CC EMBL; AF078986; AAL85491.1; -.
CC EMBL; AF078985; AAL85490.1; -.
CC EMBL; BC005808; AAK05808.1; -.
CC EMBL; BC011051; AAH11051.1; ALT_INIT.
CC Genew; HGNC:19261; MTO1.
CC InterPro; IPR001327; FAD_pyr_redox.
CC InterPro; IPR002218; GIDA.
CC InterPro; IPR004416; GIDA_sub.
CC InterPro; IPR001100; Pyr_redox.
CC Pfam; PF01134; GIDA; 1.
CC PRINTS; PR00368; FADPNR.
CC PROSITE; PS00411; PNDRTASE1.
CC PRODOM; PRO03738; GIDA; 1.
CC PROSITE; PS01280; GIDA_1; 1.
CC PROSITE; PS01281; GIDA_2; 1.
CC Alternative splicing.
CC VARSPPLIC 1 74
CC Missing (in isoform 6).
CC /FTid=VSP_001748.
CC Missing (in isoform 1).
CC /FTid=VSP_001749.
CC Missing (in isoform 2).
CC /FTid=VSP_001750.
CC Missing (in isoform 1 and isoform 5).
CC /FTid=VSP_001751.
CC H -> Q (IN REP. 1 AND 2;
CC AAL82394/AAL82395).
CC LA -> CT (IN REF. 2; AAL35894).
CC SEQUENCE 717 AA; 79963 MW; 78F84D8633BC0ED3 CRC64;
CC -----
CC Query Match 1.3%; Score 7; DB 1; Length 717;
CC Best Local Similarity 100.0%; Pred.No. 1.6e+02;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 288 ELAERLK 294
CC DB 602 ELAERLK 608
CC -----
CC RESULT 92
CC CPTM_HUMAN
CC ID CPTM_HUMAN STANDARD; PRT; 772 AA.
CC AC Q92523; Q13389; Q99655;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC DE Carnitine O-palmitoyltransferase 1, mitochondrial muscle isoform
CC (EC 2.3.1.21) (CPT I) (CPTI-M) (Carnitine palmitoyltransferase I like
CC protein).
CC GN CPTI1B.
CC OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC GO; GO:0005739; C.mitochochndrion; TAS.
CC -----
CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
CC NCB1_TaxID=9606;
CC [1]
CC SEQUENCE FROM N.A.
CC TISSUE=Heart;
CC MEDLINE=96283810; PubMed=8679700;
CC Yamazaki N., Shinohara Y., Shima A., Yamanaka Y., Terada H.;
CC "Isolation and characterization of cDNA and genomic clones encoding
CC human muscle type carnitine palmitoyltransferase I.";
CC Biochim. Biophys. Acta 1307:157-161(1996).
CC [2]
CC SEQUENCE FROM N.A.
CC Adams M.D.;
CC Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC [3]
CC SEQUENCE FROM N.A.
CC MEDLINE=98008026; PubMed=9344464;
CC Zhu H., Shi J., de Vries Y., Arvidson D.N., Cregg J.M.,
CC Woldegiorgis G.;
CC "Functional studies of yeast-expressed human heart muscle carnitine
CC palmitoyltransferase I.";
CC Arch. Biochem. Biophys. 347:53-61(1997).
CC [4]
CC SEQUENCE FROM N.A.
CC TISSUE=Heart;
CC MEDLINE=97224516; PubMed=9070950;
CC Britton C.H., Mackey D.W., Esser V., Foster D.W., Burns D.K.,
CC Yarnall D.P., Froguel P., McGarry J.D.;
CC "Fine chromosome mapping of the genes for human liver and muscle
CC carnitine palmitoyltransferase I (CPT1A and CPT1B).";
CC Genomics 40:209-211(1997).
CC [5]
CC SEQUENCE FROM N.A.
CC TISSUE=Skeletal muscle;
CC van der Leij F.R., Takens J., van der Veen A.Y., Terpstra P.,
CC Kuipers J.R.G.;
CC Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC [6]
CC SEQUENCE FROM N.A.
CC MEDLINE=97367931; PubMed=9224698;
CC Yamazaki N., Yamanaka Y., Hashimoto Y., Shinohara Y., Shima A.,
CC Terada H.;
CC "Structural features of the gene encoding human muscle type carnitine
CC palmitoyltransferase I.";
CC PNAS Lett. 409:401-406(1997).
CC -!- CATALYTIC ACTIVITY: Palmitoyl-CoA + L-carnitine = CoA + L-
CC palmitoylcarnitine.
CC -!- PATHWAY: Fatty acid beta-oxidation cycle.
CC -!- SUBCELLULAR LOCATION: Mitochondrial outer membrane.
CC -!- TISSUE SPECIFICITY: Strong expression in heart and skeletal
CC muscle. No expression in liver and kidney.
CC -!- SIMILARITY: Belongs to the carnitine/choline acetyltransferase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D87812; BAA13461.1; -.
CC EMBL; U62317; AAB03343.1; -.
CC EMBL; U66828; AAB40651.1; -.
CC EMBL; U62733; AAC51122.1; -.
CC EMBL; Y08682; CAA69938.1; -.
CC EMBL; Y08683; CAA69939.1; -.
CC EMBL; AS003286; BAA21492.1; -.
CC PIR; G02860; G02860.
CC Genew; HGNC:2329; CPT1B.
CC MIM; 601987; -.
CC GO; GO:0005739; C.mitochochndrion; TAS.
```



```

DR GO: 0004095; F:carnitine O-palmitoyltransferase activity; TAS.
DR GO: 000635; P:fatty acid beta-oxidation; TAS.
DR InterPro: IPR000542; Carn acyl trans.
DR Pfam: PF00755; Carn acyltransf_1.
DR PROSITE: PS00439; ACYLTRANSF_C_1; 1.
DR PROSITE: PS00440; ACYLTRANSF_C_2; 1.
KW Transferase; Acyltransferase; Mitochondrion; Outer membrane;
KW Fatty acid metabolism; Transport; Transmembrane; Multigene family;
KW Polymorphism.
FT DOMAIN 1 47 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 48 73 POTENTIAL.
FT DOMAIN 74 102 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT TRANSMEM 103 122 POTENTIAL.
FT DOMAIN 123 122 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 473 473 POTENTIAL.
FT VARIANT 531 531 E -> K (in dbSNP:470117).
FT VARIANT 664 664 /FTID=VAR_011739.
FT VARIANT 664 664 /FTID=VAR_011740.
SQ SEQUENCE 772 AA; 87801 MW; P7E3ED40643DFC81 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 772;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 EEKLAAL 176
DB 378 EEKLAAL 384
|||||

RESULT 93
CPTM_RAT STANDARD; PRT; 772 AA.
AC Q63704;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Carnitine O-palmitoyltransferase I, mitochondrial muscle isoform
DE (BC 2.3.1.21) (CPT I) (CPTI-M) (Carnitine palmitoyltransferase I
DE like protein).
GN CPTIB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Brown adipose tissue;
RX MEDLINE=5246880; PubMed=7729550;
RA Yamazaki N., Shinohara Y., Shima A., Terada H.;
RT "High expression of a novel carnitine palmitoyltransferase I like
RT protein in rat brown adipose tissue and heart: isolation and
RT characterization of its cDNA clone.";
RL FEBS Lett. 363:41-45(1995).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=Sprague-Dawley; TISSUE=Heart;
RX MEDLINE=96215127; PubMed=8636126;
RA Esser V., Brown N.F., Cowan A.T., Foster D.W., McGarry J.D.;
RT "Expression of a cDNA isolated from rat brown adipose tissue and
RT heart identifies the product as the muscle isoform of carnitine
RT palmitoyltransferase I (M-CPT I). M-CPT I is the predominant CPT I
RT isoform expressed in both white (epididymal) and brown adipocytes.";
RL J. Biol. Chem. 271:6972-6977(1996).
CC -1- CATALYTIC ACTIVITY: Palmitoyl-CoA + L-carnitine = CoA + L-
CC palmitoylcarnitine.
CC -1- PATHWAY: Fatty acid beta-oxidation cycle.
CC -1- SUBCELLULAR LOCATION: Mitochondrial outer membrane.
CC -1- TISSUE SPECIFICITY: High expression in heart, skeletal muscle and
CC brown adipose tissue. Also expressed in white adipose tissue, but
CC not in liver.
CC -1- SIMILARITY: Belongs to the carnitine/choline acetyltransferase
CC family.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR BMEBL: D43623; BAA07733.1; -.
DR PIR: S65532; S65532.
DR InterPro: IPR000542; Carn acyl trans.
DR Pfam: PF00755; Carn acyltransf_1.
DR PROSITE: PS00439; ACYLTRANSF_C_1; 1.
DR PROSITE: PS00440; ACYLTRANSF_C_2; 1.
KW Transferase; Acyltransferase; Mitochondrion; Outer membrane;
KW Fatty acid metabolism; Transport; Transmembrane; Multigene family.
FT DOMAIN 1 47 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 48 73 POTENTIAL.
FT DOMAIN 74 102 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT TRANSMEM 103 122 POTENTIAL.
FT DOMAIN 123 122 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 473 473 POTENTIAL.
SQ SEQUENCE 772 AA; 88216 MW; AFBB4BC67B47FA47 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 772;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 EEKLAAL 176
DB 378 EEKLAAL 384
|||||

RESULT 94
CALD_HUMAN STANDARD; PRT; 793 AA.
AC Q05682; Q13978; Q13979; Q14741; Q14742;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Caldesmon (CDM).
DE CALDI OR CDM OR CAD.
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Lung fibroblast;
RX MEDLINE=91358497; PubMed=1885618;
RA Novy R.E., Lin J.L.-C., Lin J.J.-C.;
RT "Characterization of cDNA clones encoding a human fibroblast
RT caldesmon isoform and analysis of caldesmon expression in normal and
RT transformed cells.";
RL J. Biol. Chem. 266:16917-16924(1991).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 4).
RC TISSUE=Aorta;
RX MEDLINE=92209999; PubMed=1555769;
RA Humphrey M.B., Herrera-Sosa H., Gonzalez G., Lee R., Bryan J.;
RT "Cloning of cDNAs encoding human caldesmons.";
RL Gene 112:197-204(1992).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 5).
RX MEDLINE=93101679; PubMed=1465449;
RA Hayashi K., Yano H., Hashida T., Takeuchi R., Takeda O., Asada K.,
RA Takahashi E.-I., Kato I., Sobue K.;
RT "Genomic structure of the human caldesmon gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:12122-12126(1992).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Skin;

```

RA MEDLINE-2238257; PubMed=12477932;
RA Klausner R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.J., Skalska U., Smalley D.B.,
RA Scherch A., Schain J.E., Jones S.J.M., Marra M.A.,
RA "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Actin- and myosin-binding protein implicated in the
CC regulation of actomyosin interactions in smooth muscle and
CC nonmuscle cells (could act as a bridge between myosin and actin
CC filaments). Stimulates actin binding of tropomyosin which
CC increases the stabilization of actin filament structure. In muscle
CC tissues, inhibits the actomyosin ATPase by binding to p-actin.
CC This inhibition is attenuated by calcium-calmodulin and is
CC potentiated by tropomyosin. Interacts with actin, myosin, two
CC molecules of tropomyosin and with calmodulin. Also play an
CC essential role during cellular mitosis and receptor capping.
CC -!- SUBCELLULAR LOCATION: On thin filaments in smooth muscle and on
CC stress fibers in fibroblasts (nonmuscle) (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=5;
CC Name=1; Synonyms=H-CAD;
CC IsoId=Q05682-1; Sequence=Displayed;
CC Name=2; Synonyms=WI-38 L-CAD I;
CC IsoId=Q05682-2; Sequence=VSP_004155;
CC Name=3; Synonyms=HELA L-CAD I;
CC IsoId=Q05682-3; Sequence=VSP_004154, VSP_004155;
CC Name=4; Synonyms=WI-38 L-CAD II, 1-CAD;
CC IsoId=Q05682-4; Sequence=VSP_004156;
CC Name=5; Synonyms=HELA L-CAD II;
CC IsoId=Q05682-5; Sequence=VSP_004154, VSP_004156;
CC -!- TISSUE SPECIFICITY: High-molecular-weight caldesmon (isoform 1) is
CC predominantly expressed in smooth muscles, whereas low-molecular-
CC weight caldesmon (isoforms 2, 3, 4 and 5) are widely distributed
CC in non-muscle tissues and cells. Not expressed in skeletal muscle
CC or heart.
CC -!- DOMAIN: The N-terminal part seems to be a myosin/calmodulin-
CC binding domain, and the C-terminal a tropomyosin/actin/calmodulin-
CC binding domain. These two domains are separated by a central
CC helical region in the smooth-muscle form.
CC -!- PTM: In non-muscle cells, phosphorylation by CDC2 during mitosis
CC causes caldesmon to dissociate from microfilaments.
CC Phosphorylation reduces caldesmon binding to actin, myosin, and
CC calmodulin as well as its inhibition of actomyosin ATPase
CC activity. Phosphorylation also occurs in both quiescent and
CC dividing smooth muscle cells with similar effects on the
CC interaction with actin and calmodulin and on microfilaments
CC reorganization (By similarity).
CC -!- SIMILARITY: Belongs to the caldesmon family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M64110; AAA35636.1; -

DR EMBL; M83216; AAA58420.1; -
DR EMBL; M83216; AAA58419.1; -
DR EMBL; D90452; BAA14418.1; -
DR EMBL; D90453; BAA14419.1; -
DR EMBL; BC040354; AAB40354.1; -
DR PIR; JH0628; JH0628
DR Genew; HGNC:1441; CALDL.
DR GK; Q05682; -
DR MIN; 114213; -
DR GO; GO:0005856; C:cytoskeleton; TAS.
DR GO; GO:0003779; F:actin binding; TAS.
DR GO; GO:0005516; F:calmodulin binding; TAS.
DR GO; GO:0005523; F:tropomyosin binding; TAS.
DR InterPro; IPR006017; Caldesmon.
DR InterPro; IPR006018; Caldesmon_LSP.
DR Pfam; PF02029; Caldesmon; 1.
DR PRINTS; PR01076; CALDESMON.
DR KW Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;
KW Repeat; Alternative splicing.
FT DOMAIN 319 375 3 X 14 AA TANDEM REPEATS OF B-E-E-K-R-A-
FT 1. A-E-E-R-Q-R-I-K.
FT REPEAT 319 332 1.
FT REPEAT 333 346 2.
FT REPEAT 347 360 3.
FT DOMAIN 26 207 MYOSIN AND CALMODULIN-BINDING
FT 564 621 (BY SIMILARITY).
FT DOMAIN 664 674 TROPOMYOSIN-BINDING (POTENTIAL).
FT DOMAIN 653 686 TROPOMYOSIN-BINDING (POTENTIAL).
FT DOMAIN 716 722 STRONG ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 768 793 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 39 46 WEAK ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 81 86 POLY-ARG.
FT DOMAIN 189 196 POLY-THR.
FT DOMAIN 376 379 POLY-GLU.
FT DOMAIN 540 543 POLY-GLU.
FT DOMAIN 580 583 POLY-ARG.
FT DOMAIN 597 600 POLY-GLU.
FT MOD_RES 724 724 PHOSPHORYLATION (BY CDC2)
FT MOD_RES 730 730 (BY SIMILARITY).
FT MOD_RES 753 753 PHOSPHORYLATION (BY CDC2)
FT MOD_RES 759 759 (BY SIMILARITY).
FT MOD_RES 789 789 PHOSPHORYLATION (BY CDC2)
FT MOD_RES 789 789 (BY SIMILARITY).
FT VARSPLIC 1 24 MODPERRERERKREEMLEER -> MLGGSGSHGRSL
FT AALSQ (in isoform 3 and isoform 5).
FT VFTID=VSP_004154.
FT Missing (in isoform 2 and isoform 3).
FT VFTID=VSP_004155.
FT Missing (in isoform 4 and isoform 5).
FT VFTID=VSP_004156.
FT V -> M (IN REF. 1).
FT CONFLICT 530 530 2A0DC63D16DD6B5P CRC64;
SQ SEQUENCE 793 AA; 93250 MW; 2A0DC63D16DD6B5P CRC64;
Query Match 1.3%; Score 7; DB 1; Length 793;
Best Local Similarity 100.0%; Pred. No. 1.7e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 248 REEEKR 254
DB 596 REEEKR 602
RESULT 95
SYQ_LUPLU STANDARD; PRT; 794 AA.
ID SYQ_LUPLU
AC P52780;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutaminyl-tRNA synthetase (SC 6.1.1.18) (Glutamine--tRNA ligase)
DE (GlnRS).
OS Lapsinus luteus (Yellow lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
OX NCBI_TaxID=3873;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Ventus.
RA Slatcka M., Rozek M., Barciszewski J.;
RT Isolation and characterization of a cDNA clone encoding a plant gene
of aminoacyl-tRNA synthetase.",
RL (In) Plant Gene Registry PCR95-103.
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA (Gln) = AMP +
diphosphate + L-glutaminyl-tRNA (Gln).
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X91787; CAA62901.1; --
DR PIR; T09643; T09643.
DR HSP; P00962; IGTR.
DR InterPro; IPR004514; Glns.
DR InterPro; IPR000924; Glu tRNA-synt 1c.
DR InterPro; IPR001412; tRNA-synt 1.
DR InterPro; IPR007639; tRNA-synt 1c R1.
DR InterPro; IPR007638; tRNA-synt 1c R2.
DR Pfam; PF00749; tRNA-synt 1c; 1.
DR Pfam; PF03950; tRNA-synt 1c; 1.
DR Pfam; PF04558; tRNA-synt 1c R1; 1.
DR Pfam; PF04557; tRNA-synt 1c R2; 1.
DR PRINTS; PR00987; TRNASYNTHGLU.
DR TIGRFAMS; TIGR00440; glns; 1.
DR PROSITE; PS00178; AA TENA_LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT SITE 277 287 "HIGH" REGION.
FT SITE 505 509 "KMSKS" REGION.
FT BINDING 508 508 ATP (BY SIMILARITY).
SQ SEQUENCE 794 AA; 90482 MW; 4C7D02C0795269E1 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 794;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 QLVDAKL 325
DB 175 QLVDAKL 181
|||||
RESULT 96
Y104_SYNYS
ID Y104_SYNYS STANDARD; PRT; 811 AA.
AC P54371;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein slr0104.
GN SLR0104.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,

RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO B SUBTILIS YQFF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D64004; BAAL0634.1; --
DR PIR; S76690; S76690.
DR InterPro; IPR006674; HD.
DR InterPro; IPR003607; Met phosphohydro.
DR InterPro; IPR006675; Unchar_HDIG.
DR Pfam; PF01966; HD; 1.
DR SMART; SM00471; HDC; 1.
DR TIGRFAMS; TIGR00277; HDIG; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 376 396 POTENTIAL.
FT TRANSMEM 409 429 POTENTIAL.
FT TRANSMEM 442 462 POTENTIAL.
FT TRANSMEM 486 496 POTENTIAL.
FT TRANSMEM 516 549 POTENTIAL.
FT TRANSMEM 529 549 POTENTIAL.
SQ SEQUENCE 811 AA; 90254 MW; F6015BEBF804F255 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 811;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 EKRKEV 257
DB 103 EKRKEV 109
|||||
RESULT 97
ENV_HVIND
ID ENV_HVIND STANDARD; PRT; 846 AA.
AC P18799;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Envelope glycoprotein GP160 precursor [Contains: Exterior membrane
DE glycoprotein (GP120); transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (NDK isolate) (HIV-1).
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11695;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90034200; PubMed=2806917;
RA Spire B., Sire J., Zachar V., Rey F., Barre-Sinoussi F., Galibert F.,
RA Hampe A., Chermann J.C.;
RT "Nucleotide sequence of HIV-1-NDK: a highly cytopathic strain of the
RT human immunodeficiency virus.";
RL Gene 81:275-284(1989).
CC -1- MISCELLANEOUS: NDK, ISOLATED FROM A ZAIRIAN PATIENT AFFECTED WITH
AIDS, AND IS A HIGHLY CYTOPATHOGENIC STRAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----

or send an email to license@isb-sib.ch.

```

CC -----
DR EMBL; M27323; AAA44873.1; -.
DR PIR; JQ0066; VCLJND.
DR HIV; M27323; ENV$NDK.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
KW Signal.
FT SIGNAL 1 29
FT CHAIN 30 501
FT CHAIN 502 846
FT DISULFID 53 73
FT DISULFID 118 200
FT DISULFID 125 191
FT DISULFID 130 152
FT DISULFID 213 242
FT DISULFID 223 234
FT DISULFID 291 328
FT DISULFID 374 435
FT DISULFID 381 408
FT CARBOHYD 87 87
FT CARBOHYD 129 129
FT CARBOHYD 151 151
FT CARBOHYD 179 179
FT CARBOHYD 182 182
FT CARBOHYD 229 229
FT CARBOHYD 236 236
FT CARBOHYD 257 257
FT CARBOHYD 271 271
FT CARBOHYD 284 284
FT CARBOHYD 290 290
FT CARBOHYD 351 351
FT CARBOHYD 382 382
FT CARBOHYD 388 388
FT CARBOHYD 392 392
FT CARBOHYD 395 395
FT CARBOHYD 401 401
FT CARBOHYD 438 438
FT CARBOHYD 451 451
FT CARBOHYD 452 452
FT CARBOHYD 501 501
FT CARBOHYD 506 506
FT CARBOHYD 515 515
FT CARBOHYD 527 527
SQ SEQUENCE 846 AA; 96476 MW; 8A3B9DA527DB2E83 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 846;
Best Local Similarity 100.0%; Pred. No. 1.8e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 249 EEEEXRK 255
Db 144 EEEEXRK 150

RESULT 98
ARPB YEAST
ID ARPB YEAST STANDARD; PRT; 881 AA.
AC Q12386;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Actin-like protein ARP8.
GN ARP8 OR YOR141C OR YOR3348C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.

Query Match 1.3%; Score 7; DB 1; Length 846;
Best Local Similarity 100.0%; Pred. No. 1.8e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 249 EEEEXRK 255
Db 144 EEEEXRK 150

RESULT 98
ARPB YEAST
ID ARPB YEAST STANDARD; PRT; 881 AA.
AC Q12386;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Actin-like protein ARP8.
GN ARP8 OR YOR141C OR YOR3348C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.

or send an email to license@isb-sib.ch.

MEDLINE=97344368; PubMed=9200815;
Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,
Schwager C., Paces V., Sander C., Ansorge W.;
"DNA sequencing and analysis of 130 kb from yeast chromosome XV.";
Yeast 13:655-672(1997).
[2]
GENE NAME.
MEDLINE=97435478; PubMed=9290209;
Poch O., Winsor B.;
"Who's who among the Saccharomyces cerevisiae actin-related proteins?
A classification and nomenclature proposal for a large family.";
Yeast 13:1053-1058(1997).
-1- SIMILARITY: Belongs to the actin family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X94335; CAA64058.1; -.
EMBL; Z75049; CAA99341.1; -.
PIR; S67026; S67026.
GermOnline; 143729; -.
SGD; S0005667; ARP8.
GO; GO:0005634; Cinnuleus; IDA.
InterPro; IPR004000; Actin_like.
Pfam; PF00022; actin; 1.
SMART; SM00268; ACTIN; 1.
KW Structural protein; Cytoskeleton.
FT DOMAIN 22 27 POLY-ASP.
FT SEQUENCE 881 AA; 100208 MW; 8174851B6B077A19 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 881;
Best Local Similarity 100.0%; Pred. No. 1.9e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 289 LAERLXX 295
Db 561 LAERLXX 567

RESULT 99
KLP5 SCHPO
ID KLP5 SCHPO STANDARD; PRT; 883 AA.
AC O14343;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Kinesin-like protein 5.
GN KLP5 OR SPBC2F12.13.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=21963728; PubMed=11967147;
Garcia M.A., Koonruga N., Toda T.;
"Two kinesin-like Kin I family proteins in fission yeast regulate the
establishment of metaphase and the onset of anaphase A.";
Curr. Biol. 12:610-621(2002).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

```

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leath S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Weljens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer G., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Coffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT Nature 415:871-880(2002).
CC -!- FUNCTION: Has a role in establishing metaphase during mitosis.
CC Required for chromosome segregation where it generates tension
CC during kinetochore capturing.
CC -!- SUBUNIT: Heterodimer with Klp6.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic microtubules in interphase,
CC mitotic kinetochores in metaphase and spindle midzone in anaphase
CC and telophase.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC II SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB072924; BAB69885.1; -.
CC EMBL: Z97211; CAB10160.1; -.
CC FIR: T40128; T40128.
CC HSP: P71119; 3KAR.
CC GeneDB SPombe; SPBC2P12.13; -.
CC InterPro: IPR001752; kinesin_motor.
CC Pfam: PF00225; kinesin; 1.
CC PRINTS: PR00380; KINESINHEAVY.
CC SMART: SM00129; KISC; 1.
CC PROSITE: PS00411; KINESIN MOTOR DOMAIN1; 1.
CC PROSITE: PS00667; KINESIN MOTOR DOMAIN2; 1.
CC Mitosis; Chromosome partition; Motor protein; Microtubule;
CC ATP-binding; Coiled coil.
CC FT DOMAIN 1 388 KINESIN-MOTOR (BY SIMILARITY).
CC FT DOMAIN 396 435 COILED COIL (POTENTIAL).
CC FT DOMAIN 563 588 COILED COIL (POTENTIAL).
CC FT NP BIND 144 151 ATP (POTENTIAL).
CC SQ SEQUENCE 883 AA; 99058 MW; BBA60E22DA397E8 CRC64;
CC -----
CC Query Match 1.3%; Score 7; DB 1; Length 883;
CC Best Local Similarity 100.0%; Pred. No. 1.9e+02;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 222 RSKLSSL 228
CC |||||
CC Db 514 RSKLSSL 520
CC -----
CC RESULT 100
CC ID_KBF2_CHICK STANDARD; PRT; 906 AA.
CC AC P98150;
CC DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Nuclear factor NF-kappa-B p100 subunit [Contains: Nuclear factor NF-
DE kappa-B p52 subunit].
CN NFkB2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1] _SEQUENCE FROM N.A.
RC TISSUE=Fibroblast;
RX MEDLINE=94047379; PubMed=8230480;
RA Sif S., Gilmore T.D.;
RT "NF-kappa B p100 is one of the high-molecular-weight proteins
RT complexed with the v-Rel oncoprotein in transformed chicken spleen
RT cells";
RL J. Virol. 67:7612-7617(1993).
RN [2]
RC TISSUE=SPLEEN FROM N.A.
RX MEDLINE=94171036; PubMed=7510259;
RA Ikeda T., Hirota Y., Onodera T.;
RT "Isolation of a cDNA encoding the chicken p50B/p97 (Lyt-10)
RT transcription factor";
RL Gene 138:193-196(1994).
CC -!- FUNCTION: p100 is the precursor of the p52 subunit of the nuclear
CC factor NF-kappa-B, which binds to the kappa-B consensus sequence
CC 5'-GGGNNYYCC-3', located in the enhancer region of genes involved
CC in immune response and acute phase reactions. The precursor
CC protein itself does not bind to DNA.
CC -!- SUBUNIT: Active NF-kappa-B is a heterodimer of an about 52 kDa
CC DNA-binding subunit and the weak DNA-binding subunit p65. Two
CC heterodimers might form a labile tetramer.
CC -!- SUBCELLULAR LOCATION: Nuclear, but also found in the cytoplasm in
CC an inactive form complexed to an inhibitor (I-kappa-B).
CC -!- DOMAIN: The C-terminus of p100 might be involved in cytoplasmic
CC retention, inhibition of DNA-binding by p52 homodimers, and/or
CC transcription activation (By similarity).
CC -!- MISCELLANEOUS: NF-kappa B p100 is one of the high-molecular-weight
CC proteins complexed with the v-rel oncoprotein in transformed
CC chicken spleen cells.
CC -!- SIMILARITY: Belongs to the Rel/Dorsal family.
CC -!- SIMILARITY: Contains 6 ANK repeats.
CC -!- SIMILARITY: Contains 1 death domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U00111; AAA03717.1; -.
CC EMBL: D16367; BAA03868.1; -.
CC FIR: I50404; I50404.
CC HSP: Q00653; IAKO.
CC TRANSFAC: T01929; -.
CC InterPro: IPR002110; ANK.
CC InterPro: IPR000488; Death.
CC InterPro: IPR007110; Ig-like.
CC InterPro: IPR002909; IPT_TIG.
CC InterPro: IPR000451; NF_Rel_dor.
CC InterPro: IPR008967; p53-like.
CC Pfam: PF00023; ank; 6.
CC Pfam: PF00531; death; 1.
CC Pfam: PF00554; RHD; 1.
CC Pfam: PF01833; TIG; 1.
CC PRINTS: PR00057; NFKBTNSCPFCT.
CC SMART: SM00248; ANK; 6.
CC SMART: SM00005; DEATH; 1.

DR SMART; SM00429; IPT; 1.
 DR PROSITE; PS00088; ANK_REPEAT; 5.
 DR PROSITE; PS00297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS0017; DEATH_DOMAIN; FALSE_NEG.
 DR PROSITE; PS01204; REL_1; 1.
 DR PROSITE; PS0254; REL_2; 1.
 KW DNA-binding; Transcription regulation; Activator; Nuclear protein;
 KW Repeat; ANK repeat; Phosphorylation.
 FT DOMAIN 37 342 REL-LIKE [RHD].
 REPEAT 472 501 ANK 1.
 REPEAT 511 540 ANK 2.
 REPEAT 544 573 ANK 3.
 REPEAT 582 611 ANK 4.
 REPEAT 616 646 ANK 5.
 REPEAT 650 679 ANK 6.
 DOMAIN 771 857 DEATH.
 FT DOMAIN 336 340 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 DOMAIN 349 396 GLY-RICH.
 FT CONFLICT 55 55 V -> G (IN REF. 2).
 FT CONFLICT 272 272 S -> F (IN REF. 2).
 FT CONFLICT 337 337 N -> K (IN REF. 2).
 FT CONFLICT 395 395 C -> G (IN REF. 2).
 FT CONFLICT 409 409 R -> SG (IN REF. 2).
 FT CONFLICT 430 430 R -> A (IN REF. 2).
 FT CONFLICT 533 533 E -> Q (IN REF. 2).
 FT CONFLICT 569 569 A -> G (IN REF. 2).
 FT CONFLICT 653 653 T -> A (IN REF. 2).
 FT CONFLICT 690 693 VRVP -> SEA (IN REF. 2).
 FT CONFLICT 736 736 L -> RC (IN REF. 2).
 FT CONFLICT 759 774 SPILSCPPPPSRNHL -> RPDTELTTPRAGNV (IN REF. 2).
 FT CONFLICT 779 779 T -> S (IN REF. 2).
 FT CONFLICT 792 792 Y -> D (IN REF. 2).
 FT CONFLICT 816 816 I -> D (IN REF. 2).
 FT CONFLICT 821 823 ASA -> SVSL (IN REF. 2).
 FT CONFLICT 830 830 P -> A (IN REF. 2).
 FT CONFLICT 885 885 R -> A (IN REF. 2).
 SQ SEQUENCE 906 AA; 99667 MW; 0E0CE20DB4F30B62 CRC64;
 Query Match 1.3%; Score 7; DB 1; Length 906;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 229 CRELQKH 235
 Db 423 CRELQKH 429
 Search completed: June 7, 2004, 14:45:41
 Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OK protein - protein search, using sw model

Run on: June 7, 2004, 14:42:11 ; Search time 19 Seconds
(without alignments)
2683.237 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 530

Sequence: 1 KSSFGQPEAGPEGAQERPSQ.....APSTEASQGTPQEPISARA 530

Scoring table:

OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 300 summaries

Database :

PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	3-2	676	2 JC7222	77K muscle-derived
2	9	1-7	594	2 S50611	hypothetical prote
3	9	1-7	1402	2 I46707	translation initia
4	8	1-5	162	2 T29728	hypothetical prote
5	8	1-5	211	2 T32354	hypothetical prote
6	8	1-5	224	2 T52558	hypothetical prote
7	8	1-5	229	2 S87236	translation elonga
8	8	1-5	280	2 S71315	probable membrane
9	8	1-5	281	2 C88638	deoxyguanosine kin
10	8	1-5	312	2 T08985	protein F58P6.1 (l
11	8	1-5	450	2 E71909	hypothetical prote
12	8	1-5	467	2 D8710	hypothetical prote
13	8	1-5	500	2 H97769	protein C43G2.1 (l
14	8	1-5	525	2 H75514	hypothetical prote
15	8	1-5	531	2 A55887	glutamyl-tRNA synt
16	8	1-5	535	2 C36811	caldesmon, non-mus
17	8	1-5	535	2 A37994	Rf1 protein - saim
18	8	1-5	568	2 E90364	hypothetical prote
19	8	1-5	594	2 S62141	transcription init
20	8	1-5	758	2 S65169	hypothetical prote
21	8	1-5	761	2 T32183	hypothetical prote
22	8	1-5	928	2 T52292	endopeptidase Clp
23	8	1-5	952	2 T52456	endopeptidase Clp
24	8	1-5	952	2 T49283	AtClpC - Arabidops
25	8	1-5	1100	2 AE3243	conjugal transfer
26	8	1-5	1101	2 T03419	traA protein - Agr
27	8	1-5	1902	2 C97702	cell surface antig
28	8	1-5	2052	2 T18519	myosin X - bovine
29	7	1-3	23	2 S45032	homeotic protein S

30	7	1.3	60	2	B91253	hypothetical prote
31	7	1.3	62	2	AG1905	hypothetical prote
32	7	1.3	75	2	AC1755	hypothetical prote
33	7	1.3	79	2	D75326	hypothetical prote
34	7	1.3	105	2	S78290	ribosomal protein
35	7	1.3	120	2	S10587	cystatin C - rat
36	7	1.3	121	2	T44473	conserved hypotet
37	7	1.3	127	2	S07085	cystatin C precurs
38	7	1.3	128	2	T44497	hypothetical prote
39	7	1.3	133	2	S77948	major allergen Par
40	7	1.3	136	2	B65240	hypothetical prote
41	7	1.3	136	2	D64964	hypothetical prote
42	7	1.3	138	2	S52933	major allergen Par
43	7	1.3	138	2	A90258	conserved hypotet
44	7	1.3	154	2	G98112	hypothetical prote
45	7	1.3	155	2	B95248	hypothetical prote
46	7	1.3	158	2	T49567	related to attachm
47	7	1.3	159	2	T32043	hypothetical prote
48	7	1.3	163	2	E71183	hypothetical prote
49	7	1.3	172	2	A69454	hypothetical prote
50	7	1.3	176	2	AE2913	conserved hypotet
51	7	1.3	177	2	T37212	hypothetical prote
52	7	1.3	179	2	C85165	hypothetical prote
53	7	1.3	180	2	B33445	H+-transporting tr
54	7	1.3	181	2	E70209	conserved hypotet
55	7	1.3	187	2	E95846	probable ABC trans
56	7	1.3	190	2	T45013	hypothetical prote
57	7	1.3	193	2	A81149	outer membrane lip
58	7	1.3	193	2	H81874	probable outer mem
59	7	1.3	195	1	SAVLH1	delta large antige
60	7	1.3	195	1	SAVLH1	delta large antige
61	7	1.3	195	1	SAVLDM	delta large antige
62	7	1.3	195	2	S53112	delta large antige
63	7	1.3	199	2	G75028	delta antigen - he
64	7	1.3	201	2	G95852	H+-transporting Ar
65	7	1.3	211	2	C69375	conserved hypotet
66	7	1.3	214	1	SAVLDM	hypothetical prote
67	7	1.3	214	2	A53175	delta large antige
68	7	1.3	214	2	JC1062	delta large antige
69	7	1.3	218	2	S73675	delta large antige
70	7	1.3	225	2	T05514	hypothetical prote
71	7	1.3	229	1	PQEP82	hypothetical prote
72	7	1.3	229	2	D90831	antiterminator Q -
73	7	1.3	229	2	C90912	probable antitermi
74	7	1.3	229	2	F90972	antiterminator (im
75	7	1.3	229	2	G85688	probable antitermi
76	7	1.3	229	2	C85820	antiterminator (im
77	7	1.3	232	2	G30292	hypothetical prote
78	7	1.3	232	2	E97347	probable transcrip
79	7	1.3	235	2	A86150	TIN6.24 protein -
80	7	1.3	237	2	H97687	hypothetical 17.5k
81	7	1.3	237	2	B95374	hypothetical prote
82	7	1.3	240	2	F90885	hypothetical prote
83	7	1.3	240	2	A85733	hypothetical prote
84	7	1.3	240	2	S84897	hypothetical prote
85	7	1.3	246	2	H72732	hypothetical prote
86	7	1.3	248	1	JH0252	myelin P0 protein
87	7	1.3	253	2	G72598	probable ABC trans
88	7	1.3	262	2	T04809	hypothetical prote
89	7	1.3	271	2	T37222	hypothetical prote
90	7	1.3	277	2	JC6142	probable secreted
91	7	1.3	280	2	AI0182	deoxyguanosine kin
92	7	1.3	285	2	C86423	carboxylesterase (
93	7	1.3	297	2	P64750	unknown protein, i
94	7	1.3	306	2	T21938	regulatory protein
95	7	1.3	308	2	G84701	hypothetical prote
96	7	1.3	319	2	D64303	hypothetical prote
97	7	1.3	322	2	T45568	hypothetical prote
98	7	1.3	326	1	E64547	hypothetical prote
99	7	1.3	326	2	B71960	nitrogen fixation
100	7	1.3	327	2	B82918	hypothetical prote
101	7	1.3	327	2	C90278	DNA-directed RNA p
102	7	1.3	332	2	F81743	hypothetical prote
						conserved hypotet


```

; APPLICANT: Millet, Isabelle
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: MacDougall, John R.
; APPLICANT: Mishra, Vishnu
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glenda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Ort, Tatiana
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Zernuoen, Bryan D.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-533C
; CURRENT APPLICATION NUMBER: US/10/336,472
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,681
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 16
; LENGTH: 1137
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-336-472-16

Query Match      8.8%; Score 236.5; DB 12; Length 1137;
Best Local Similarity 24.6%; Pred. No. 6e-06;
Matches 147; Conservative 80; Mismatches 196; Indels 175; Gaps 26;

QY 12 EGAQERPSQAAPAYEAEGFGSSQAPRKPGEQAQARTAGSALRDVSELSQLEDILSTYC 71
DB 237 EGSLEQEKVEMDLE-----RAKRLGDLKLQTQSS--IMDLND-KQQLERLKKD 286
QY 72 VDNNGGPGDGAQCEPAEPEDA-----EKSRTYVAR-----NGPEPTPVYGEKEP 119
DB 287 FSLN-----ALNARIEDQALGSQIQKGLKELOARIBELEFEELBAERTAKVKKLR 338
QY 120 SKGDPNTEIRQS-DEVG-----DRHRPQKKKAKGLGKEITLLMQLTNTLSTPEE 171
DB 339 SLSRELEBEISERLEEAGGATSVQIEMNKKRAEFQWRDLSEATL-----CHEA 389
QY 172 KLAALCKKY-----BELLEHNSOKMKLQKQSQVLQEKDHLRGHSHKAVLARSKLES 227
DB 390 TAAALRKKHADSVAEGLGQIDNLQVRQKLEKSEFKLELDVDTSNNEQIIRAKANLEK 449

```

```

228 LCR-----ELQRHN-----RSLKE----- 241
DB 450 MCKTLEDQWNEHRSKAEETQRSVNDLTQRAKLOTENGELSRLQDEKEALISQLTRGKIT 509
QY 242 -----EGVQARAREEKEKVTSH-POVTLNDIQLOMEQHNER-----NSK 281
DB 510 YTOQLEUDUKRQLEEBVKAKNALAHALQARHDCDILRQYEESTAKAELQRLVLSKANSE 569
QY 282 LR-----QENMELAERLKKLIBQYELRESHIDKV-----FKHK-----DLOQ 318
DB 570 VAQWRTKYETDAIORTTELEEAQKLAQKQLOBAEAVEAVNAKCSLSKTKHRLQNEIED 629
QY 319 QLVDAKLOQAQEMKAEERHOREKDFLLKAEVSORMCELMKQOETHLQKQALATYTKF 378
DB 630 LAMVDVERNA-----AAAALDKKQORNFKILAE--WKQYEESSOSELESSOKARSLSLELF 584
QY 379 EEFONTLSKSSSEVFTTFQKEMEKMTKKI-----KGLKEXETMYR 417
DB 685 -KLKNAYEESLEHLETFKREKNKLOBEISDLTEQLGSSGKTHLELKVQKQLEAKQNELQ 743
QY 418 SWHSSNKALLEMAEKKVVRDELEGLQVKIQRLKGC-----RALQTERNDLNKRVQD 471
DB 744 SALEEA-EASLEHHEGKILR-AQLEFNQIKABIERKLAEKDEMEQAKNEL--RVVD 797

RESULT 43
US-10-408-765A-254
; Sequence 254, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 254
; LENGTH: 3225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-254

Query Match      8.8%; Score 236.5; DB 16; Length 3225;
Best Local Similarity 22.8%; Pred. No. 2.2e-05;
Matches 119; Conservative 99; Mismatches 178; Indels 125; Gaps 18;

QY 52 LRDVSELSQLEDILSTYCVDNNGGPGDGAQCEPAEPEDAERSRTYVARNGSPEPTP 111
DB 2248 LQENKELLSQLESTPHLY-----HSSQELAKLESELKSL-----KDQLTD 2389
QY 112 VYGEKPSKGPDPNTEETIRQSDVEGDRDHRPQ-----KKAKGLGKEITLLMOT 162
DB 2290 LNSLKECKEKGKNGLEGIIRQOEADIQSKFSVEQLETDQASRELTSLRHEENKQK 2349
QY 163 LNTLSTPEKLAALCKYAELEHNSOKMKLQKQSQVLQEKDHLRGHSHKAVLAR 222
DB 2350 IISLSGKSE--AIQVAIAELQQH---DKRIKELNLLSQEENIVLEENKAV--- 2401
QY 223 SKLESICRLQRHNSLKEEGVQARAEEBEKEKVT-----HFQVT 264
DB 2402 DKTNQLMETHL-----KTIKKENIQQAQLDSFVKSSSLQNDRDRIVDYQLEERHLSII 2457
QY 265 LNDIQLOMEQHHERNSKLQKQ-----NMELAEKLLKLIBQYELREBEHIDKVPKH 313
DB 2458 LEKDQLIQAAAAA--NNKLEKRIGRSHRSHMDLNSXAKLDABELI-QY-----REDLNQVITI 2512

```


Db 1492 YEEISDLETLKRENNLQEQEISDLTEQIAEGGKRIHELEKIKKQVEQEKSELQAALSEA 1551
Qy 445 QVKIORLEKLCALQTERNDLKRIV 469
Db 1552 EASLEHEEGKILRIQLELNQVAKSEV 1576

RESULT 46
US-10-104-047-3636
; Sequence 3636, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3636
; LENGTH: 650
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3636

Query Match 8.7%; Score 234.5; DB 15; Length 650;
Best Local Similarity 21.6%; Pred. No. 4e-06;
Matches 126; Conservative 100; Mismatches 257; Indels 99; Gaps 19;

Qy 7 PEAGPGAQERPSQAAPVAAEAGPGSQAPRPEGAQARTAGSALRDVSEB-----LS 60
Db 58 PETTSGGCHSP-----EQTQNRALQKEEKAKSHQHQAALREIEAQQHTIRILT 108
Qy 61 RQLEDILSTYCVDNNGGPGEDGAQGEPAEPEDAEKSRITYVARNGEP---EPTPVYVGRK 117
Db 109 CQKTEETALYYGDAARFEDGNLTGTFPSFNALSQ---AFRGSPLGCVSTSLIPGS 164
Qy 118 EPSKGDPN-----TEHQRDSQVGDHRRPQKKAAGLKGKEITLLMQTL-----NTL 166
Db 165 KDLAHLHSHWHPAGSLQALSASVSTRH-----KKADRYIEELTKERDALSLELYRNTI 218
Qy 167 STPEEKLALCKVAELLEEHRNSQMKLLQKQSQLVQEKDHLRG--EHSKAVILARSK 224
Db 219 TNEELK-----KNREL-----QKLPASEKSEIQLNVKELKRLERAKFLPQVQ 266
Qy 225 LESLCRELQRHNSLKEEGVORAREBEKKEKVTSHFQVTLNDIQLOMEQHNERNKSLRQ 284
Db 267 TNLQEMMRQEBELRQ-EKKIRKQEKMRQEBELRQEGGKMRQEEKMRQEKRLRE 325
Qy 285 ENMELAEKRLKLEOVELREHIDKVPKHKDLOQLVDKLOQAQEMLKEAERHOREKD 344
Db 326 QEKLEQEKELREKRLRQ-----EQMQRQ--EEKMRQEBKMRQEEKMRQEB 376
Qy 345 FLKAEVRSQRMCELMKQEQTHLQKQALYTERPEEP--QNTLSKSEVFTTPK----- 396
Db 377 RLWEQEKMRQEQEQGQDQERNWEQERLREKMRQEKMRQEKMRQEKMRQEQE 436
Qy 397 -----QEMKMTKIKGL-EKETT-----YRSRWSSNK-----ALLENA 431
Db 437 KKTQDQEKMRQEBIRERBCKMRQEBEETVREQEKMRQEKMRQEKMRQEQEQLRPEQ 496
Qy 432 EEXTVRKELEGLQVKQLRLEKLCALQTERNDLKRIVQDLSAGGQSLTDSQPER---R 488
Db 497 XEKLWEQEKMRQEBEKIWEQEKLRDQEMWQEKMRQEKMRQEKMRQEKMRQEBEERIR 556
Qy 489 PEGGAQAPSPPVTEAPCPGAPSTASQGTGQPEPTSARA 530
Db 557 KEMQERLPEHEBRCSEPCPLPPSKVLCNMGHTSVSEPAGGEA 598

RESULT 47
US-10-336-472-20
; Sequence 20, Application US/10336472
; Publication No. US20040043929A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.
; APPLICANT: Ballinger, Robert A.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Berghs, Constance
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E.
; APPLICANT: Millet, Isabelle
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Pasturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: MacDougall, John R.
; APPLICANT: Mishra, Vishnu
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Smithson, Glennnda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Ort, Tatiana
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Zernhusen, Bryan D.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-533C
; CURRENT APPLICATION NUMBER: US/10/336,472
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,681
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 20

```

; LENGTH: 1859
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-336-472-20

Query Match      8.7%; Score 234; DB 12; Length 1859;
Best Local Similarity 23.4%; Pred. No. 1.6e-05;
Matches 117; Conservative 101; Mismatches 175; Indels 108; Gaps 24;

QY 8 EAGPEGAGQERPSQAAPAV-----EAGPGSSQAPRPEGAQARTAGSALRDVSELS 60
Db 1363 KANSEVAQWRKYETDAIQRTTELEEEAKKLAQRLQBAEEAVEAVNAKSSLEKTKRLQ 1422
QY 61 ROLEDILSTYCVDNNGGPGEDGAGGEPAPEDAKSRITYVARNGEPTPTVYGEKPS 120
Db 1423 NEIEDLM-----VDVERSNAAAAAL-----DKQR 1447
QY 121 KGPDPNTEIROSDEVGDRDRHRRPOEKKKAGLGKBITLIM-----OTLNTLSTPBKLAAL 176
Db 1448 NFDKILAEWKQYBESQSELESSQ--KEARSLSLSTELFKLKNAYBESLEHLETFKRENK 1505
QY 177 CKKYAELLEHRNSQOMKLLQKQSQVLVQKDHRLG-----EHSKAVLASKLE- 226
Db 1506 QBEISDLTEQLGSSGKTIHELEKVKQLAEKMLQSALEBAEASLEHEGKILRAQLEF 1565
QY 227 -SLCRELQHRNSLKEBVGQVQARPEEKEKVTSHFQVTTINDIQLQME--CHNERNSKLRV 284
Db 1566 NQIKABIER-----KLAEKDSEMEQAKENHLRV--VDSLQTSLDABETSRNEALRV 1614
QY 285 ENKVELAERLKLIEQVELRE--EHIDKVPKHKDL---QQQLVDAKLQAEQMLKEA--BE 337
Db 1615 KKKQ-TERSKLAEO-ELIETSERVQLHSQNTSLNKKNDADLSLQTFEVEAVQBC 1672
QY 338 RHQREKDFLLKAVESQRCM--ELMKQOET--HLKQALALYTEKPEFONILSKSSEV-F 392
Db 1673 RNABEK---AKRAITDANMAEELKEQTSIHLERKQKMGQTTIKOLQRLDEAEQIAL 1729
QY 393 TTFQKQEMRMTKKIKLEKETTMYRSRWBSNKALEMAEETKVRDK-LEGLOVKIQRL 451
Db 1730 KGGKQKQLEARVRELENE-----LE-AEQK--RNAESVKGMRKSERRI 1771
QY 452 EKLCRALQTERNDLNKRVODL 472
Db 1772 KELTYQTEEDRNKL-LRUQDL 1791

RESULT 48
US-10-336-472-22
; Sequence 22, Application US/10336472
; Publication No.: US20040043929A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.
; APPLICANT: Ballinger, Robert A.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Berghs, Constance
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E.
; APPLICANT: Millet, Isabelle
; APPLICANT: Padigaru, Muralidhara
```

```

; APPLICANT: Patturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: MacDougall, John R.
; APPLICANT: Mishra, Vishnu
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Smithson, Glennda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Ort, Tatiana
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Zehrusen, Bryan D.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-533C
; CURRENT APPLICATION NUMBER: US/10/336,472
; CURRENT FILING DATE: 2003-01-03
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,681
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 22
; LENGTH: 1935
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-336-472-22

Query Match      8.7%; Score 234; DB 12; Length 1935;
Best Local Similarity 22.4%; Pred. No. 1.6e-05;
Matches 127; Conservative 100; Mismatches 177; Indels 162; Gaps 25;

QY 8 EAGPEGAGQERPSQAAPAV-----EAGPGSSQAPRPEGAQARTAGSALRDVSELS 60
Db 1363 KANSEVAQWRKYETDAIQRTTELEEEAKKLAQRLQBAEEAVEAVNAKSSLEKTKRLQ 1422
QY 61 ROLEDILSTYCVDNNGGPGEDGAGGEPAPEDAKSRITYVARNGEPTPTVYGEKPS 120
Db 1423 NEIEDLM-----VDVERSNAAAAAL-----DKQR 1447
QY 121 KGPDPNTEIROSDEVGDRDRHRRPOEKKKAGLGKBITLIM-----OTLNTLSTPBKLAAL 176
Db 1448 NFDKILAEWKQYBESQSELESSQ--KEARSLSLSTELFKLKNAYBESLEHLETFKRENK 1505
QY 177 CKKYAELLEHRNSQOMKLLQKQSQVLVQKDHRLG-----EHSKAVLASKLE- 226
Db 1506 QBEISDLTEQLGSSGKTIHELEKVKQLAEKMLQSALEBAEASLEHEGKILRAQLEF 1565
QY 227 -----SLCRELQHRNSL-----KEEQ----- 243
```

Db 1566 NQIAKIERKAEKDEMEQAKRNHLRVVDSLSQTSIDAETSRNEALRVKQKQEGDLNEM 1625
Qy 244 -VORA---REEEKREVTSHEQVTLNDIQLQMEQHNRNSKLRQNMELAER----- 292
Db 1626 EIQLSHANMAEAQKVKS-IQSLKQTOIQUDDAVRANDDLK-ENTAIVERRNNILQA 1683
Qy 293 ----LKKLJEQVE---LREHIDKVFHKDL-----OQQLVDAKLOQAQEMLKE 334
Db 1684 ELEELRAVVQETSRKLAQELIETSERVQLLHSQNTSLNQKQKQADLSQLQTEVEE 1743
Qy 335 A--EERHOREKDFLLKEAVESQSMC--ELMKOQET--HLKQQLALYTEKFEFQNTLSK 388
Db 1744 AVQECNAHEK---AKGATDAAWAEELKQDSHAHLERKQKQMEQTIKQLQRLDEA 1800
Qy 389 SEV-FTTFQEMEKMTKIKLEKETMYRWSWESSKALLEMAEKTVRDKE-LEGLOV 446
Db 1801 EQIALKGGKKQLQKLEAVRELENE-----LE-AEQK--RNAESVKQMEK 1842
Qy 447 KIQRLKLCALQOTERDNLKRVODL 472
Db 1843 SERRIKELTYQTEEDRNKL-LRLQDL 1867

RESULT 49

US-10-108-260A-4080
; Sequence 4080, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4080
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4080

Query Match 8.6%; Score 233.5; DB 15; Length 512;
Best Local Similarity 24.7%; Pred. No. 3.5e-06;
Matches 115; Conservative 72; Mismatches 176; Indels 103; Gaps 17;

Qy 75 NQGG-----PGEDGACGEPAEPDAEKSRITYVARNGEPEPTPVYGEKPSKG 122
Db 5 NQGGESADGHI SCPPPSIIGNAGEKSLSDAKKK-----KS 42
Qy 123 DPNTBEIRQSDVGDHRR-----POEKKAAGLQKE--ITLLMOTLNTLSTPEKLA 175
Db 43 NRKEDDVWASGTV--KRLHKTSGECERKTKKSLSKEDLIQLLSIMEGELQAREDV 100
Qy 176 LC--KYVAELLBEHRNSQOMKLLQ--KKOSQLVQEKDHLRGHSHKAVLARSKL 231
Db 101 LKTEKTPVELBAHYGSAEPKVLVRHDAIIAQKSGIDYVFKPISELDELBEKQKE 160
Qy 232 LQR----HNRSLKEGVQARAEHEEKKEVTSFQVTLNDIQLQMEQHNRNSKLR 285
Db 161 TYRMLBOLLAAKCHERTVYELENEKHKHTDYMNKSDDFNTLLEQERERLKLLEQ 220
Qy 286 ----NMELERLKLLEQ-----YELREEHIDKVFHKDLQQLVD--AKLO 326
Db 221 YQARKEKENAKRLNKLDELVLKSPALMLVDERQMGHEQL-----GLOSQKV 276
Qy 327 QAQEMLKEASERHOREKDFILKEAV-----ESQRMCLMKQOETH---LQOOLA 372
Db 277 EEEELKAITSKSKEDQKLLKLVDFEHKASRFSQHEBMAKLANQESHNRQLAKLV 336
Qy 373 LYTEKPEFQNT-----LSKSSVFTTFQEMEKMTKIKLEKETMYRWSWESSK 429
Db 337 GLTORIEELETNKNQKAE-----ELQELRDLKIAKGCQGNSSLMVAEVLNRKRV 389

Qy 430 MA---BEKTVRDKLEGLQVKIQLEKLCALQOTERDNLKRVODL 472
Db 390 MEGKDEBITTKTESQCRBLRKLKQLEEEHSHKFEFRLEVEKLOCRMSEL 435
RESULT 50
US-10-080-608A-11
; Sequence 11, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-608A-11

Query Match 8.5%; Score 230.5; DB 14; Length 3878;
Best Local Similarity 18.4%; Pred. No. 6.3e-05;
Matches 119; Conservative 110; Mismatches 195; Indels 223; Gaps 23;

Qy 31 GSSQAPRPEGAQARTAGSGALRDVSELSQLIEDIISTYCVDNNO-----GGPGSDG 83
Db 161 GAQDSPTHELMSELA--GKQHEI--EELNSELSEMRVYTGTEGLQQQOEPEAAIKORDG 217
Qy 84 -----AQGEPAEPEDAESKRTYVARNGEPEPTPVYGEKPSKGDPT-- 126
Db 218 IITQLTANLQARREKQETMEFLELTQSQKLIQIFQOLQASETLRNSHSTSTAADLLQ 277
Qy 127 ---BEIROSVDGRDHRPQEKKAGLGEITLLMOTLNTLSTPEKLAALCK-- 179
Db 278 AKQILTHQQOLEEQCHLLEDYQKKEDFTWQISFLQEKIKYEMEQDKGVNSNKESIQ 337
Qy 180 -----YAEELLSEHRNSOKOM-----KLQKQSQLVQEKDHLRGHSHKAVLARS 223
Db 338 EKFTIIEBLNTKIIEEKKTLLEKDKLTATKLLGELQEQIVQKNQELK--NMKLELTNS 395
Qy 224 K-----LESCLRELQRNR---SLKBEQVQARAEHEEKKE-- 256
Db 396 KQKERSQSEIEIKQLMGTVLELQKRNHKSQPETDIVQRMEOETORLQRLAEELDEMYGQ 455
Qy 257 -----VTSH-----FQVTLNDIQLQME 273
Db 456 QIVQMKQELIRQHMMAQMEEMKTRHKGEMENALRSYSNTVNEHQIKLNVAINELNKLQ 515
Qy 274 QINERNSKLR-----QENMELAEELKCLTEQVELREEHIDKVPK-----HK 314
Db 516 DTNSQKELKEELGHLILEKCALQRLDELVEELSFSEQIQARQRTIAEOBSKLEAAHK 575
Qy 315 DLOQ--QLVDAKLOQAQEMLKEAERH-----QREKDFLLKEAVESORM-- 356
Db 576 SLSTVEDLKAETVSASERKLELEHAEVNTYKIKLEMLKEKKNVLAORMAESQAELE 635
Qy 357 -----CELMAKQOE-----THLKOOL-----ALYTEKF 378
Db 636 RLRTQLLSHEEELSCLKLELEIHRINIEKLNGLGHIYKQIIDGLQWENKSKLETMOF 695
Qy 379 EEFQNTLSKSSVFP-----TTFQEMEKMTKIKLEKETMYRWSWESSK 427
Db 696 EK--DNLITKQNLILEISKLDLOOSLVNSKSEMTQLINELQKEITELR----- 744
Qy 428 LMASEKTVRDKLEGLQVKIQLEKLCALQOTERDNLKRVODLSA 474
Db 745 -QEEKKGTLEQVEVQLQKTELLBK---QMEKENDLQEKFAQLEA 787

Search completed: June 7, 2004, 14:44:21
Job time : 135 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:22:20 ; Search time 46 Seconds
(without alignments)
1108.294 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 2702

Sequence: 1 KSSPQPPAGPBGAGQERSQ.....APSTEASGTGPQEPSARA 530

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

PIR 78:**

1: Piri:**

2: Piri2:**

3: Piri3:**

4: Piri4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1233	45.6	676	2 JC7222	77K muscle-derived
2	375	13.9	346	2 F87844	protein T22C1.6 [i
3	372.5	13.8	335	2 T25110	hypothetical prote
4	269	10.0	992	2 T46337	hypothetical prote
5	263.5	9.8	892	2 T50985	related to transcr
6	261.5	9.7	1407	1 S28589	trichohyalin - rab
7	259	9.6	886	2 H69378	conserved hypothet
8	259	9.6	955	2 S24348	myosin heavy chain
9	256	9.5	936	2 S39083	myosin heavy chain
10	256	9.5	1938	1 JX0178	myosin heavy chain
11	256	9.5	2677	2 A38194	desmoplakin I - hu
12	253	9.4	1206	2 T34021	protein kinase SK2
13	252.5	9.3	880	2 P75103	conserved hypothet
14	251.5	9.3	1349	1 A46691	trichohyalin - she
15	249.5	9.2	1233	2 T14157	serine/threonine p
16	249.5	9.2	2139	2 T18296	myosin heavy chain
17	249	9.2	1392	2 A43336	microtubule-vesicl
18	249	9.2	1938	2 A52993	skeletal myosin he
19	247	9.1	1001	2 T17365	serine/threonine p
20	246	9.1	1427	2 S22695	restin - human
21	245	9.1	1377	2 T18055	myosin heavy chain
22	245	9.1	1940	2 A28320	myosin heavy chain
23	243.5	9.0	993	2 S49461	synaptosomal compl
24	242	9.0	1534	2 A56734	ribosome receptor,
25	242	9.0	3259	1 A56539	giantin - human
26	240	8.9	1931	2 A59234	slow myosin heavy
27	240	8.9	2663	1 S28261	centromere protein
28	239.5	8.9	1898	1 A45973	trichohyalin - hum
29	239	8.8	4574	2 G02520	plectin - human

30	239	8.8	4684	2 A59404	plectin [imported]
31	238.5	8.8	4687	1 A39638	plectin - rat
32	237.5	8.8	1690	2 T13030	microtubule bindin
33	237	8.8	1934	2 T48153	myosin heavy chain
34	236.5	8.8	3225	2 T52300	giantin - human
35	236	8.7	978	2 A70387	conserved hypothet
36	236	8.7	1410	1 A57013	early endosome ant
37	234.5	8.7	1231	2 T18532	serine/threonine pr
38	234	8.7	1935	1 A37102	myosin beta heavy
39	234	8.7	1935	2 A59286	myosin heavy chain
40	233.5	8.6	845	2 T48176	synaptosomal compl
41	233.5	8.6	1939	1 A46762	myosin alpha heavy
42	232.5	8.6	1432	2 B85431	trichohyalin like
43	232	8.6	875	2 A23767	myosin heavy chain
44	232	8.6	1935	1 S06006	myosin beta heavy
45	231.5	8.6	1927	2 A59236	embryonic muscle m
46	231.5	8.6	1938	1 S06005	myosin alpha heavy
47	231.5	8.6	1938	2 T49464	alpha cardiac myos
48	231	8.5	1388	2 S74245	serine/threonine-s
49	230.5	8.5	1642	2 T08880	NMDA receptor-bind
50	230.5	8.5	1939	2 T48175	myosin heavy chain
51	230.5	8.5	3187	2 JC5837	364K Golgi complex
52	229	8.5	1269	2 F84730	probable myosin he
53	229	8.5	1976	2 A59252	myosin heavy chain
54	229	8.5	2442	2 T08621	centrosome associa
55	228	8.4	1164	2 T24806	hypothetical prote
56	228	8.4	1388	2 S70633	serine/threonine-s
57	228	8.4	1938	1 A40997	myosin heavy chain
58	227.5	8.4	1744	2 F86151	PI003.10 protein -
59	226.5	8.4	1790	2 S67593	transport protein
60	226	8.4	1999	1 S21801	myosin heavy chain
61	225.5	8.3	1909	2 A45592	liver stage antige
62	225.5	8.3	1939	2 T18372	repeat organellar
63	225.5	8.3	1940	1 A24922	myosin heavy chain
64	225	8.3	746	2 T47237	myosin II heavy ch
65	224.5	8.3	911	2 S51441	hypothetical prote
66	224.5	8.3	1115	2 T41342	probable coiled-co
67	223.5	8.3	1133	2 T22976	hypothetical prote
68	223	8.3	492	2 A28616	M5 protein precurs
69	223	8.3	764	2 T51302	myosin heavy chain
70	223	8.3	1974	2 T30010	hypothetical prote
71	222.5	8.2	1940	1 S04090	myosin heavy chain
72	222.5	8.2	2057	2 S61477	myosin II heavy ch
73	222	8.2	2101	2 A42184	nuclear mitotic ap
74	221.5	8.2	946	2 S28061	SCP1 protein - rat
75	221.5	8.2	1039	2 S18199	myosin heavy chain
76	221	8.2	520	2 S35575	myosin heavy chain
77	221	8.2	1091	2 T34107	hypothetical prote
78	221	8.2	1961	1 A61231	myosin heavy chain
79	220.5	8.2	2007	1 B43402	myosin heavy chain
80	220.5	8.2	2017	1 A36014	myosin heavy chain
81	220	8.1	741	2 S39082	myosin heavy chain
82	220	8.1	924	2 S06117	myosin heavy chain
83	220	8.1	2094	2 S33124	tpi protein - huma
84	219.5	8.1	852	2 D72230	conserved hypothet
85	219.5	8.1	2253	2 T30336	nuclear/mitotic ap
86	219	8.1	1804	2 T34518	myr 6, unconventional
87	219	8.1	1846	2 A59289	nestin - golden ha
88	218.5	8.1	1738	2 T14867	interaptin - slime
89	218.5	8.1	1964	2 A59282	nonmuscle myosin I
90	218.5	8.1	2168	2 T30171	ninein - mouse
91	218	8.1	725	2 A47168	cardiac morphogene
92	218	8.1	1017	2 PC4035	cell-cycle-depende
93	218	8.1	1155	2 B70356	chromosome assembl
94	218	8.1	1313	2 A48467	myosin heavy chain
95	217.5	8.0	967	1 A35075	kinesin heavy chai
96	217.5	8.0	1956	2 T16416	hypothetical prote
97	217.5	8.0	1972	1 A41604	myosin heavy chain
98	217.5	8.0	1979	1 S03166	myosin heavy chain
99	216	8.0	555	2 C96667	unknown protein, 7
100	216	8.0	1251	2 A56677	neuronal cell cycl
101	216	8.0	1940	2 A59287	myosin heavy chain
102	215.5	8.0	1034	2 T32297	hypothetical prote

103	215.5	8.0	1354	2	S69211	serine/threonine-s
104	215.5	8.0	1959	1	A33977	myosin heavy chain
105	215	8.0	484	2	S46489	M1 protein precurs
106	215	8.0	697	2	T07111	MAE binding filame
107	215	8.0	1837	2	T41023	probable nuclear p
108	214.5	7.9	2245	2	T18278	myosin heavy chain
109	214	7.9	1300	2	I53799	CG1 protein - huma
110	214	7.9	1354	2	S74244	serine/threonine-s
111	214	7.9	1356	2	S32763	kinectin1 - human
112	214	7.9	1957	2	T38077	hypothetical coile
113	213	7.9	1875	2	S38173	myosin-like protei
114	213	7.9	1992	2	A47297	myosin heavy chain
115	212.5	7.9	1496	2	T056334	hypothetical prote
116	212	7.8	1300	2	T18364	ro-3 protein - Neu
117	212	7.8	2954	2	T14156	kinasin-related pr
118	211.5	7.8	2116	2	A26655	myosin heavy chain
119	211	7.8	1290	2	A55094	chromosomal protei
120	210.5	7.8	863	2	S37040	paramyosin - tapew
121	210.5	7.8	1313	2	F96673	hypothetical prote
122	210.5	7.8	1597	2	S68420	citrion - mouse
123	210	7.8	451	2	G70241	hypothetical prote
124	210	7.8	1827	2	T16370	hypothetical prote
125	209.5	7.8	729	2	T50989	hypothetical prote
126	209.5	7.8	1039	2	S62509	probable vesicular
127	209.5	7.8	1044	2	T50213	probable vesicular
128	209	7.7	476	2	B96667	hypothetical prote
129	209	7.7	1175	2	C35815	myosin heavy chain
130	209	7.7	1201	2	A35815	myosin heavy chain
131	209	7.7	1201	2	B35815	myosin heavy chain
132	209	7.7	1270	2	T09194	adaptor protein in
133	209	7.7	1325	2	T42722	male-enhanced anti
134	209	7.7	1702	2	T14050	protein kinase (BC
135	209	7.7	2385	2	A32491	myosin heavy chain
136	209	7.7	2411	2	B32491	myosin heavy chain
137	208.5	7.7	652	2	B59102	hypothetical prote
138	208.5	7.7	1046	2	T42734	cytoplasmic linker
139	208	7.7	853	2	T51505	hypothetical prote
140	208	7.7	1029	2	H96658	hypothetical prote
141	208	7.7	1475	2	T33318	hypothetical prote
142	208	7.7	1968	1	S05697	myosin heavy chain
143	207.5	7.7	621	2	K10450	myosin heavy chain
144	207.5	7.7	1938	1	PMKX1	myosin heavy chain
145	207	7.7	734	2	T27055	hypothetical prote
146	207	7.7	1426	2	T00337	hypothetical prote
147	207	7.7	1818	1	S73852	hypothetical prote
148	206.5	7.6	583	2	C84788	probable myosin he
149	206.5	7.6	690	2	S41009	hypothetical prote
150	206.5	7.6	705	2	B88566	protein T05G5_9 [i

ALIGNMENTS

```

RESULT 1
JC7222      77K muscle-derived protein - chicken
C:Species: Gallus gallus (chicken)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C:Accession: JC7222
R:Uyeda, A.; Fukui, I.; Fujimori, K.; Kiyosue, K.; Nishimune, H.; Kasai, M.; Taguchi, T.
Biochem. Biophys. Res. Commun. 269, 564-569, 2000
A:Title: Mbp77: A novel neurite-outgrowth-promoting protein predominantly expressed in
A:Reference number: JC7222; MUID:20175243; PMID:10708594
A:Accession: JC7222
A:Molecule type: mRNA
A:Residues: 1-676 <UYE>
A:Cross-references: GS:089999; NID:G7619883; PIDN:BAA94755.1; PID:G7619884
A:Experimental source: crus muscle
C:Comment: This protein, a glycoprotein and a neurite-outgrowth-promoting protein, is im
C:Keywords: coiled coil; glycoprotein; leucine zipper; muscle

Query Match      45.6%  Score 1233; DB 2; Length 676;
Query Local Similarity 53.9%  Pred. No. 5.8e-45;

```

Matches 255, Conservative 84, Mismatches 100; Indels 34; Gaps 7

```
Qy 2 SSPGQBPAGPEGAGQRPSQAAPAVEAGPSSQAPRXPBGQAQARTAGGALRDVSEELSR 61
Db 15 TSPTQDNQG-----QSKAEPVPVPSQP-SP-TNQTSAQPEKATC-----DISEELNR 59
Qy 62 QLEDILSTYCVDDNNOGQPGDGA---QGRPAEPEDAEEKSNTYVARNGCEPSPTPVVGEEK 117
Db 60 QLEDILITTY-----GSAASLVEKEGTTAETDKPEKEDVGSMDACEDVNEESKD 110
Qy 118 EPKSGEDPNTBEIROSDVGRDHRRPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALC 177
Db 111 KPAPGDAS----RAKEPSASKEQK--LEKKILKGLGKEATLLMQSLNKLTTPEEKLDLEL 164
Qy 178 KTYAELLBEHRNSOKMKLQKQSQSLVQSKDHLRGHSHS KAVLARSKLESICRELQHRNR 237
Db 165 KTYAELLBEHRAEQKQLKYLQKQQAQITKEKQDLQSEHSRAILARSKLESICRELQHRNK 224
Qy 238 SLKEEGVQVAREBEERKEVNTSHFQVTLNDITQLQMEQHNRNRSKLROENMELARLKKLI 297
Db 225 TLKEETLQAREBDEKKKEITNHFGGTLSLQAIQQOSHRNMKLCQENTELAEKLSII 284
Qy 298 EQVELREEHIDKVPKHKDLQQLQVDAKLQAQAEMLKEAEERHOREBKDFLLAKEAVESQMC 357
Db 285 DQVELREEHLDKIPKHELQKLVDAKLQEQSQEMWKEAEERHOREKKEVYLLNQAAEWLQA 344
Qy 358 ELMKQOETHLQKQALVYTEKPEEPQNTLSKSSVFYTFKQEMEKMTKKIKLEKETTMYR 417
Db 345 KMLKEQNTVLQAQITLYSERFEFQKTLTSNEVFATPKQEMEKMTKKMKKLEKDTATWK 404
Qy 418 SRVSSNKALLEMAEETKVRDKELEGQVKIQRLEKLCRALQTERNDLNKRVQ 470
Db 405 SRFENCRRALLDMEEKAMRTKEYECFVKLQRLNLCRALQERNELYRKIK 457

RESULT 2
F87844
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: F87844
R:anonymous, The C. elegans Sequencing Consortium.
Science 283, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating
A:Reference number: A75000; MUTD:99089613; PMID:19851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Proj
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103,
A:Accession: F87844
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <STO>
A:Cross-references: GS:chr_I; PIDN:CAA99923.1; PID:G3880023; GSPDB:GN00019; CES
C:Genetics:
A:Map position: 1
```

Db 181 YMEYKTKDILKILKTAKLESASIQVKS-----GMEKDELAKIMLEETARVGGALKTEK 236
QY 366 HLKQQLALYTKPEEFQNTLSSESSEVFTFKQEMKMTKIKKLEKETTMYRSWESSNK 425
Db 237 ALREQVQYSAKYSILTSCLSNSEAFDFKDEISRVNKKCMQVEKGLSYKKSDEANK 296
QY 426 ALL-----EMAERTVDRKLEGLQVQIKRLEKLCRALQ 459
Db 297 KVLVLTMTNQEVAEKIATSDK-----KIQMLENLCRALR 330

RESULT 3

T25110
hypothetical protein T22C1.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R:McMurray, A.
A:Accession: T25110
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1335 <WIL>
A:Cross-references: EMBL:Z75550; PIDN:CAA99923.2; GSPDB:GN00019; CESP:T22C1.6
A:Experimental source: clone T22C1
C:Genetics:
A:Map position: 1
A:Introns: 6/1; 46/3; 66/3; 91/3; 285/2

Query Match 13.8%; Score 372.5; DB 2; Length 335;
Best Local Similarity 30.8%; Pred. No. 3.3e-09;
Matches 103; Conservative 70; Mismatches 120; Indels 41; Gaps 9;

QY 152 LGKEI--TLIMOTLNTLSTPE--EKLALCKYVAELLEHRNSQKMKLQKQSQVLQV 208
Db 1 MGNFDEALIKSLE--GYFDAKQVLIKLALE--SEKQAEIKIKLVLDYKVKVVDL 56
QY 209 DHLRGHSHKAVL-----AKSLKESLCRELQRNRSLSKEGVQARBEERKKEVTSHFQV 263
Db 57 TEKLENNQILLRTBEAKSKLEELCRGLQKXANQHTREACAKMKKLEVERGLAVEQLKV 116
QY 264 TLNDLQLOMEQHNRNSKLROENMELAEKLIKLEQYE-----LREEHIDKVPK 312
Db 117 TLKDLEKTMWAGRSKSDSLAENDKKLSKPSFSGHQYEEMKVIDQIQKKEKYWESYCK 176
QY 313 HKDLQQLVDKLAQACMLKEAEERHOREKDFLLKEAVESQRMCELMKQOETHLKQOLA 372
Db 177 TKDLETKLITAKLESASIQVKS-----GMEKDELAKIMLEETARVGGALKTEKALREQVQ 232
QY 373 LYTEKPEEFQNTLSKSSEVFTFKQEMKMTKIKKLEKETTMYRSWESSNKALL----- 428
Db 233 EYSAKYSILTSCLSNSEAFDFKDEISRVNKKCMQVEKGLSYKKSDEANKKVLVLT 292
QY 429 ---EMAERTVDRKLEGLQVQIKRLEKLCRALQ 459
Db 293 TNQEYAEKATSDK-----KIQMLENLCRALR 319

RESULT 4

T46337
hypothetical protein DKFZp43402413.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46337
R:Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23037
A:Cross-references: EMBL:AL389891; GSPDB:GN00116; NCSP:B7F18.40
C:Genetics:
A:Status: preliminary
A:Molecule type: mRNA

A:Residues: 1-992 <AAA>
A:Cross-references: EMBL:AL137265
A:Experimental source: adult testis; clone DKFZp43402413
C:Genetics:
A:Note: DKFZp43402413.1

Query Match 10.0%; Score 269; DB 2; Length 992;
Best Local Similarity 21.9%; Pred. No. 0.00022;
Matches 133; Conservative 102; Mismatches 208; Indels 164; Gaps 20;

QY 7 PEAGPEGAQERPSQAAPAEVRECP-----GSSQAPRKEPGAQARTAQSGALRDYSEELSR 61
Db 104 EKADPTG-----SRPAKASEKEAPEEDVDAGEEGSRREAAKEPKKASALEGSSDSAQ 158
QY 62 QLE-----DILSTY-----CYDNNQGG 78
Db 159 ELEISEHMKBPQLSDSIASDPKSPHGLDFGFRSRISEHLLDVLSPVLGACACQAQOPL 218
QY 79 PGEDGAQGEPAEPZ-DAKSRITYVARGEP-----BPTPVYVGEKPSKGDPTNEEI 129
Db 219 GIEDKDSQSSQDELSQSKGLEERLSPPLPHEERAQSPRSLATEEPPEQSGPEQEW 278
QY 130 QSDVEVGD-----RDHRRPQEKKA 149
Db 279 KEARELGDSASLSLQSLQREQAQPPACEKGEQHSQAESLGPQOEAEDEPKVA 338
QY 150 KGLGKEITLLMOTLNTLSTPE-----KLAALCKYVAELLEHRNSQKMKLQKQSQVL 205
Db 339 VSPTPPVSPVSTPEVAPPQLSEAAALKAMEEAVQVLE-----QDQHLLSKQETMQ 393
QY 206 QSKDHLRGHSHKAVL--RSKLSLCSRELQRNRSLSKEGVQARBEERK-----RKVT 258
Db 394 QREKLCQOEETELRLHQKEQSLSSLRERLQKAEIEEA-RMREESQRLSWLRAQVQ 452
QY 259 SHFQVTLNDILOME-----QHNRNSKLROENMELAEKLIKLEQVYELREH 306
Db 453 SSTQADEQDQIRAEQBASLOKLEBESQKAEASLEQKNQMLEQLKEEIEASEKSEA 512
QY 307 IDKVPKHLQLOQ---OLVDKLAQQAEMKE-----AEERHOREKDFLLKEAVE 352
Db 513 ALNAAKALQQLREQLGEGERKEAVATLEKEHSALEKLCSSLEAKHREVVSSLOKQIOE 572
QY 353 SQRMCELMKQ---QETHLKQALALYTEKFEFFQNTL--SKSSEVFTFKQEMKMTKIK 406
Db 573 AQCKEQAQKLCQVREHVHQSVHVAGYEHLSLLREKQVSEGEHERLDM----- 628
QY 407 KKLKETTMYRSWESSNKALLEMAEKTVDRKLEGLQVQIKRLEKLC-RALQTERNDL 465
Db 629 -KEHQQVMAKAREQYB-----AEERKQAEILLGHLTGELRLQRAHRELETVROBQ 680
QY 466 NKRVDOL 472
Db 681 HKRLIEDL 687

RESULT 5

T50985
related to transcription factor TMF [imported] - Neurospora crassa
N:Alternate names: protein B7F18.40
C:Species: Neurospora crassa
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50985
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, July 2000
A:Reference number: Z25286
A:Accession: T50985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-892 <SCH>
A:Cross-references: EMBL:AL389891; GSPDB:GN00116; NCSP:B7F18.40
A:Experimental source: BAC clone B7F18; strain OR74A
C:Genetics:
A:Gene: NCSP:B7F18.40

A:Map position: 6
A:Introns: 63/1; 551/3

Query Match 9.8%; Score 263.5; DB 2; Length 892;
Best Local Similarity 23.3%; Pred. No. 0.00033;
Matches 140; Conservative 101; Mismatches 240; Indels 121; Gaps 27;
QY 1 KSSPGQPEAGPEGAQ---ERPQ---QAAPAVAEAGSGSQAPRKPEGA--QAATAQSGA 51
DB 101 QASPRQMDAPSASTDSIERPNFADKASPNVSSPRASADTPRKQSDTTQPPVAPLDS 160
QY 52 LRQVSELSQLBILSTYCVNNQGGPDEGAQGPAPBPAEDAEKRTTVARNGEPEPTP 111
DB 161 ADDIKEVEAGVSE-----KADSRPSTD--QPSFNTEDSAPT-----TESKLEAPA 206
QY 112 VYGEKEPSGDNTEIRQSDSVGRDHRPQEKKKAKGLGKEITLLMOTL-----163
DB 207 EILAKTEPKSVVTEKAPDGGKSDTPQNDQEIYAYVERIDALAKQYLAREAS 266
QY 164 -----NTLSTP-----BEKLAALCKKYAELLEHN--SOKQ-----MKLQKQSOVLQV 207
DB 267 AAARKEALSAPAGSAEKKLAEDQQAQIAQMEGKNLASNEQKLTILKNLKKQADEKD 326
QY 208 KDLRGHSHKAVLARSKLSLGR---ELQHNLSLKEEGVORABEEEEKKXVTSHEQVT 264
DB 327 MGNLKAKEKADREIENLRGRARHAELEKSNELQRLDQSOEYLYLRPEVSK-DTI 385
QY 265 LNDIQLQMEQHNER-----NSKLQENMELAEKLEKLEQYELREBHIDKVPKHDL 316
DB 386 IAEIRSQIKATEQADVMSAKANDKAREQDQ---RRIAEI-----BESVEALKIEKNL 435
QY 317 QQQLVDKLOQAQEMLEKEERQREKDF-----LLKEAVESQR-----355
DB 436 ---MADRKAQADELKEAEKASEKAKALELEKAEVHMVESKLEAMVTRAEASGVGTG 492
QY 356 --MCELMKQOETHLKOQALYATEKFBFQNTLSKSEVFTTFKQEMKMTKKIKLEKET 413
DB 493 DSQAKLQVET--LQSQYSTASENWQGIETTL--RSRIVNLEKERDE-----ALQRES 542
QY 414 TWYRSRWESSKALLEMAEKTVRDKEGLQVIOLEKLCR-----ALQTERNDLNKRV 469
DB 543 DMERKAREAE--LEBAKTLQNPQDVESYRSQDLSKKRAEAEBAALAEARADFEKQ 599
QY 470 QDLSAGQSGSLTDSGPPRRPPEGCAQAPSPRYTEAPCPGAP--STEASGQGPQEPPTS 527
DB 600 QAMEA--EKELIKEEREDLQSQ-----NRPR-SWLQGLPGPFLXNEGSGPQSLSTA 652
QY 528 AR 529
DB 653 QR 654

RESULT 6
S28589
trichohyalin - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C:Accession: S28589
R:Fietz, M.J.; Rogers, G.E.
submitted to the EMBL Data Library, December 1992
A:Description: Examination of the gene encoding rabbit trichohyalin.
A:Reference number: S28589
A:Molecule type: DNA
A:Residues: 1-1407 <FIE>
A:Cross-references: EMBL:Z19092; NID:g1746; PIDN:CAA79519.1; PID:g1747
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covalent modifications to this protein include conversion of arginine to citrulline and
C:Genetics:
A:Introns: 46/3
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 9.7%; Score 261.5; DB 1; Length 1407;
Best Local Similarity 23.5%; Pred. No. 0.00065;
Matches 117; Conservative 110; Mismatches 168; Indels 103; Gaps 22;
QY 26 EABGPGSS-----QAPRKPEGAQARTAQSGALADVSEELSRQLEDILSYCYVDNQQG 78
DB 514 QEQRFQGTWRQQLQEAQRRHTLYAKFGQEQELRE--EEELQ-----EKROE 561
QY 79 PQEDGAQCEPAEPDAEKSRVTVARNGEPEPTPVVYGEKPSKGDPTNTEIRQSDVGR 138
DB 562 REREYREKELQREDEKER-----RQERERQYSELELQREQLRDR 604
QY 139 DHRPQEKKKAGKGLKTYLLMOTLNTLSTPEKLAALCKKYAELLER-HNNSQKQMKLL 197
DB 605 KLRBEQQLQER---BEELRRQERERKLRBEQ---LRQEQELQREKRLREBEQ 659
QY 198 OKQSQOLVQKDH-LRGHSHKAVLARSKLSLCLRELQHNLSLKEEGVORABEE--EK 253
DB 660 RREBQELQREERKLRBEQ---LQREERELRQERARKLRBEQQLRQEQELQER 716
QY 254 RKEVTSHPQVTLNDIQLQMEQHNERNSKLQENMEL-----AERLKKLEQYELREBHIDK 309
DB 717 ERKLRBEQQLRREQLLRQ---ERDRKLRBEQQLQESERELRRQERQQLRRER-DR 772
QY 310 VPKHD-----LQQLVDKLOQAQEMLEKEA-----ERHQRKDFLLKEA 350
DB 773 KFRBEQQLQREERELRQERKLRBEQQLQREERELRRQERARKLRBEQQLQER 832
QY 351 VESQSMC-----ELMKQOETHLKOQALYATEKFBFQNTLSKSEVFTTFKQ 398
DB 833 --EELRLRQERERKLRBEQQLRQEQELQREBA---RKLREBEQQLRQEQ---ELROE 885
QY 399 MEKMTKKIKL-----EKETTWYRSWESSKALLEMAEKTVRDKEGLQVIOLEK 454
DB 886 RDRKLRBEQQLRQEQELQREDRKLRBEQQLQESERELRRQERER---KLREBEQ 942
QY 455 CRALQTE-RNDLNKRVQD 471
DB 943 LRREBQELRRERARKLRE 960

RESULT 7
H69378
conserved hypothetical protein AF1032 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Feb-2001
C:Accession: H69378
R:Klenk, R.P.; Clayton, R.A.; Tomb, J.P.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.P.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A59250; MUID:98049343; PMID:9389475
A:Accession: H69378
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-886 <KLE>
A:Cross-references: GB:AE001032; GB:AE000782; NID:G2689355; PIDN:AAB90211.1; PID:G26495
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 9.6%; Score 259; DB 2; Length 886;
Best Local Similarity 24.8%; Pred. No. 0.00051;
Matches 113; Conservative 86; Mismatches 111; Indels 146; Gaps 22;
QY 116 EXERKSGDPTNTEIRQSDVGRDHRPQEKKKAKGLGKEITLLMOTLNTLSTPEKLA 175
DB 171 EXERLK-----EFLSQEQI-----KQKEEKA-----EIERISEEIKSLREKLS- 214
QY 176 LCKKYAELLEHNRSQKMKLQKQSGIVQSKHLRGHSHKAVLARSKLSLCLRELQ 235

Db 215 -----EVRNLESLKELBEHKSRL-----ESLRKQSSVUQVRGLBEKLELE-- 259
Qy 236 NRSLEKGVQARBEERKEV-----TSHFQVTLND----- 267
Db 260 -KQLKV-VERIBOLEKAKEVKEPKAERYSLKLLSEINQALADVEKREGDLTREA 317
Qy 268 --IOLQEHNRSLKROENMELAEELKLLILOYE-----IR 303
Db 318 AGIOAQKKAEDNSKL-EETKRIEELEREFKSHRLLETLPKPMORMQGIKAKLE 376
Qy 304 EEHf--DKVFHKLQLOQVDAKLQQAQEMLEKAEERHOR---EKDFL-----LKEAVE 352
Db 377 EKNLTPDKVEQWYDL-----LSKAKEBEKEITEKLLKLLAKKSSLKTRGAQLKKA 428
Qy 353 -----SQRCMLMKQO--ETHLKQALALYTEKFEFQNTLSKSSSVFTTFQKMEKMTYKI 406
Db 429 ELKSAERTCPVCGRELDDEHRKNIMAYETTEMKRIABELAKADIEKKLERLSKVE-- 485
Qy 407 KKLKERTMYRSRW-----EGSNKALLEMAEKTVRDKLEGL--QVKI-- 448
Db 486 KALEKQETVLKYRWVDELKALENELSHDAKLSAESSEYRKVER-LDGLRQCKILL 544
Qy 449 -----QRLSKLCRALQTERNDLNKRVOD 471
Db 545 SSASRIKELXSLREIBELAKNVESEKELRKIRE 580

RESULT 8
S24348
myosin heavy chain, embryonic and adult skeletal muscle (clone Cemb2) - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 03-Feb-1994 #sequence_revision 06-Sep-1996 #text_change 13-Feb-1998
C:Accession: S24348
R:Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
J. Mol. Biol. 225, 1143-1151, 1992
A:Title: Analysis of the chicken fast myosin heavy chain family. Localization of isoform
A:Reference number: S24348; MUID:92309413; PMID:1377278
A:Accession: S24348
A:Molecule type: mRNA
A:Status: nucleic acid sequence not shown
A:Residues: 1-955 <MOO>
A:Cross-references: EMBL:M74085
A:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: ATP; coiled coil; muscle contraction; skeletal muscle; thick filament

Query Match 9.6%; Score 259; DB 2; Length 955;
Best Local Similarity 22.6%; Pred. No. 0.00055;
Matches 118; Conservative 82; Mismatches 181; Indels 142; Gaps 16;
Qy 33 SOAPKPEGAQARTAQSGALRDVSELSRQLEDILSTYCYDNNQGGPGEAGQEPAPPE 92
Db 107 SQIOSKIEDQALGMQ---LQKKIKELQARIEL-----ESEEIERTSRA 149
Qy 93 DAERSRTYVARNGEPEPTPVVYGEKPSKGPDPNTEIRQSDEVDGDRHRFPQKCAKGL 152
Db 150 KAEKHRADLSLEEE-----ISELEAGGATATQI-----DMNKEAEAFQWRD 196
Qy 153 GKEITLMTLNTLSTPEKLAALCKKY-----AELEHNRNSQKQMLKQKQSLVQEK 208
Db 197 LEEATL-----QHEATAAALRKHADSTAEGLQIDNLQVRVKQLEKESKELMCEI 247
Qy 209 DHLRGESHKAVLARSKLESICREL-----QRNRSLEKEGVQVQAR----- 248
Db 248 DLLASNMSVSKAVANLEKWCRTLEDQLSKSEKEHQRMINDLSTQRLARLTQESGEYS 307
Qy 249 -----BEEKKEVTSH-FQVTLNDIQLQMEQHN 276
Db 308 QVBEKDALISQSRGKAFTQOIIEELKRLHEEIKAKNALAHALQSAHDCDILLREQYE 367
Qy 277 ER-----NSKL-----QENMELAEELKLLILOYELREHIDKVP 311
Db 368 SEQAKGELQALSKANSEVAQWTKYETDAIORTTELEBEAKKLAQLQDABEHV---- 423

Qy 312 KHKDLQOQLVDKALQQAQEMLEKAEERHORQKDFLLKAEAVESQRCMLKQOETHLQOQL 371
Db 424 -----EAVNAKCA-----LEKTKQRLQNEVEDLMIDVERANAACARLDKQKQNFDKIL 472
Qy 372 ALYTEKFEFQNTLSKSSSVFTTFQKMEKMTYKIKKLEKETMYRSRVSSNKALLEMA 431
Db 473 AEKWKQYETQAELEASQKESRSLSTELFAM-----KNAYEESLHLETLKRNKNLQOEI 528
Qy 432 SEKTVRDKLEGLQVKIQKLEKLCRALQTERNDLNKRVODLSA 474
Db 529 SDLT--EQIAEGGKA-IHELEKVKQIEQEKSEIQAALEAEAA 568

RESULT 9
S39083
myosin heavy chain, neonatal [similarity] - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 20-Jun-2000
C:Accession: S39083; S24350; A26821
R:Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
Submitted to the EMBL Data Library, August 1991
A:Description: Analysis of the chicken fast myosin heavy chain family. Localization of
A:Reference number: S39081
A:Accession: S39083
A:Molecule type: mRNA
A:Residues: 1-936 <MOO1>
A:Cross-references: EMBL:M74087
R:Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
J. Mol. Biol. 225, 1143-1151, 1992
A:Title: Analysis of the chicken fast myosin heavy chain family. Localization of isoform
A:Reference number: S24348; MUID:92309413; PMID:1377278
A:Accession: S24350
A:Molecule type: mRNA
A:Residues: 3-466, 'Q', 468-641, 'R', 643-936 <MOO2>
A:Cross-references: EMBL:M74087
R:Moraricity, D.M.; Barringer, K.J.; Dodgson, J.B.; Richter, H.E.; Young, R.B.
DNA 6, 91-99, 1987
A:Title: Genomic clones encoding chicken myosin heavy-chain genes.
A:Reference number: A26821; MUID:87217964; PMID:3034534
A:Accession: A26821
A:Molecule type: DNA
A:Residues: 'F', 856-936 <MOR>
A:Cross-references: GB:M16557; NID:9212371; PIDN:AAA48970.1; PID:9212372
C:Genetics: 886/3
A:Introns: 886/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; muscle; skeletal muscle

Query Match 9.5%; Score 256; DB 2; Length 936;
Best Local Similarity 22.0%; Pred. No. 0.00072;
Matches 122; Conservative 80; Mismatches 180; Indels 172; Gaps 18;
Qy 19 SQAAPAVEAGFGSSQAPRPEGQAQARTAQSGALRDVSELSRQLEDILSTYCYDNNQGG 78
Db 89 SQIOSKIEDQALGMQKQKIKELQART-----EELEEEIE----- 124
Qy 79 PUEGAGQEPAPPEDAKSRITYVARNGEPEPTPVVYGEKPSKGPDPNTEIRQSDEVG 138
Db 125 -----AERTSRKAEKGRADLSLEEE-----ISELEAGGATATQI-----DM 164
Qy 139 DHRPQEKKAKGLGKEITLMTLNTLSTPEKLAALCKKY-----AELEHNRNSQKQ 194
Db 165 NKKREAEFCWMDLDEEATL-----QHEATAAALRKHADSTAEGLQIDNLQVRK 215
Qy 195 KLLQKQSQVLQVQKHLRQEHKAVLARSKLESICREL-----QRNRSLEKEG 243
Db 216 QKLEKESKELMCEIIDLASNMESVSKASLEKTCRALEDQMSKTEKKEHQRMINDVN 275
Qy 244 VQAR-----BEEKKEVTSH-FQ 262
Db 276 AQARLTQESGEYSQVBEKDALISQSRGKAFTQOIIEELKRLHEEIKAKNALAHGLQ 335
Qy 263 VTINDIQLQMEQHN-----NSKL-----QENMELAEELKLLI 297

Db 336 SARHDCDLLREQVEEQEAKGELQALSKANSEVAQNRKTYETDAIQRTTELEAKKLA 395
QY 298 EOVLEREHIDKVPFKHDLQOOLQADKLOQAQEMLEKABERHOREKDFLLKEAVESQEMC 357
Db 396 QRLQDABEHV-----EAVNSKAS-----LETKORLQNVEDLMDIVRANSAC 440
QY 358 ELMKQOQTHLKQALALYTKFPEPQNTL-----SKSEVF---TTFKQ---EMEKMT 403
Db 441 AALDKKQKQKILSEFKQKVEYEQABLEASQKESRLSTELFKMKNAYEBSLDHLETLK 500
QY 404 KTKKLEKETMYKSRWESSNKALLEMAEKTVRDKELEGLOVKIQPLE-----KLC 455
Db 501 RENKQOQESLDLTFQABGKKAHELEKVKQIQEQKSLQASLEAREASLEHBECKIL 560
QY 456 RALQTERNDLNRV 469
Db 561 R-LQLELNQVNSEI 573

RESULT 10
JX0178
myosin heavy chain, fast skeletal muscle, adult [validated] - chicken
N:Contains: myosin ATPase [EC 3.6.4.1]
C:Species: Gallus gallus (chicken)
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text change 19-Apr-2002
C:Accession: PX0050; PX0051; PX0052; JX0178; A26365; S02082; PW0009; S39081; S24351; S05
R:Hayashida, M.; Maita, T.; Matsuda, G.
J. Biochem. 110, 54-59, 1991
A:Title: The primary structure of skeletal muscle myosin heavy chain: I. Sequence of the
A:Reference number: PX0050; MUID:92041767; PMID:1939027
A:Accession: PX0050
A:Molecule type: protein
A:Residues: 1-205 <RAY>
R:Komine, Y.; Maita, T.; Matsuda, G.
J. Biochem. 110, 60-67, 1991
A:Title: The primary structure of skeletal muscle myosin heavy chain: II. Sequence of the
A:Reference number: PX0051; MUID:92041768; PMID:1939028
A:Accession: PX0051
A:Molecule type: protein
A:Residues: 206-636 <KOW>
R:Maita, T.; Miyaniishi, T.; Matsuzono, K.; Tanioka, Y.; Matsuda, G.
J. Biochem. 110, 68-74, 1991
A:Title: The primary structure of skeletal muscle myosin heavy chain: III. Sequence of the
A:Reference number: PX0052; MUID:92041769; PMID:1939029
A:Accession: PX0052
A:Molecule type: protein
A:Residues: 201-213,632-837 <MAI>
R:Maita, T.; Yajima, E.; Negata, S.; Miyaniishi, T.; Nakayama, S.; Matsuda, G.
J. Biochem. 110, 75-87, 1991
A:Title: The primary structure of skeletal muscle myosin heavy chain: IV. Sequence of the
A:Reference number: JX0178; MUID:92041770; PMID:1939030
A:Accession: JX0178
A:Molecule type: protein
A:Residues: 833-1938 <MA2>
R:Maita, T.; Hayashida, M.; Tanioka, Y.; Komine, Y.; Matsuda, G.
Proc. Natl. Acad. Sci. U.S.A. 84, 416-420, 1987
A:Title: The primary structure of the myosin head.
A:Reference number: A26365; MUID:87092420; PMID:3467365
A:Accession: A26365
A:Molecule type: protein
A:Residues: 1-129, 'X', 131-139, 141-550, 'X', 552-754, 'X', 756-784, 'QL', 787-804, 806-810 <MA3>
R:Watanabe, B.
Biol. Chem. Hoppe-Seyler 370, 55-61, 1989
A:Title: Amino-acid sequence of the hinge region in chicken myosin subfragment-2.
A:Reference number: S02082; MUID:89228549; PMID:2713098
A:Accession: S02082
A:Molecule type: protein
A:Residues: 1144-1270 <WAT>
R:Yajima, E.
Nagasaki Igakkai Zasshi 65, 409-420, 1990
A:Title: Study on tail region of skeletal muscle myosin; primary structure and protease
A:Reference number: PW0009

A:Accession: PW0009
A:Molecule type: protein
A:Residues: 1304-1938 <YAT>
R:Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
submitted to the EMBL Data Library, August 1991
A:Description: Analysis of the chicken fast myosin heavy chain family: Localization of
A:Reference number: S39081
A:Accession: S39081
A:Molecule type: mRNA
A:Residues: 1081-1203, 'DV', 1206-1342, 'E', 1344-1544, 'S', 1546-1795, 'HV', 1798-1829, 'S', 183
A:Cross-references: EMBL:M74084
R:Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
J. Mol. Biol. 225, 1143-1151, 1992
A:Title: Analysis of the chicken fast myosin heavy chain family. Localization of isoform
A:Reference number: S24348; MUID:92309413; PMID:1377278
A:Accession: S24351
A:Molecule type: mRNA
A:Residues: 1082-1182, 'T', 1184-1203, 'DV', 1206-1342, 'E', 1344-1544, 'S', 1546-1780, 'M', 1782
A:Cross-references: EMBL:M74084
R:Watanabe, B.
Biol. Chem. Hoppe-Seyler 370, 1027-1034, 1989
A:Title: Complete amino-acid sequence of subfragment-2 in adult chicken skeletal muscle
A:Reference number: S05515; MUID:90121764; PMID:2610940
A:Accession: S05515
A:Molecule type: protein
A:Residues: 842-906, 'Q', 908-1270 <WA3>
R:Watanabe, B.
Biol. Chem. Hoppe-Seyler 370, 549-558, 1989
A:Title: Amino-acid sequence of the short subfragment-2 in adult chicken skeletal muscle
A:Reference number: S04501; MUID:89374803; PMID:2775482
A:Accession: S04501
A:Molecule type: protein
A:Residues: 852-906, 'Q', 908-1108 <WA2>
R:Matsuda, G.; Maita, T.; Miyaniishi, T.; Hayashida, M.
J. Protein Chem. 6, 33-46, 1987
A:Title: Structure and function of muscle myosin.
A:Reference number: A60877
A:Accession: A60877
A:Molecule type: protein
A:Residues: 1-139, 141-205 <MA1>
R:Gulick, J.; Kropp, K.; Robbins, J.
J. Biol. Chem. 260, 14513-14520, 1985
A:Title: The structure of two fast-white myosin heavy chain promoters. A comparative st
A:Reference number: A92507; MUID:8603956; PMID:2997212
A:Accession: A24124
A:Molecule type: DNA
A:Residues: 'M', 1-168 <GUL>
A:Cross-references: GB:M13512; GB:M12083; GB:M13510; NID:G212363; PIDN:AAA48966.1; PID:
R:Kropp, K.; Gulick, J.; Robbins, J.
J. Biol. Chem. 261, 6613-6618, 1986
A:Title: A canonical sequence organization at the 5'-end of the myosin heavy chain gene
A:Reference number: A92587; MUID:86196091; PMID:3009465
A:Accession: C25217
A:Molecule type: DNA
A:Residues: 'M', 1-56, 'T', 58-76, 'I', 78-168 <KRO>
A:Cross-references: GB:M13515; GB:M13511; NID:G212373; PIDN:AAA48971.1; PID:G555468
C:Comment: This is a fragment of the globular head.
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: acetylated amino end; actin binding; ATP; coiled coil; hydrolase; methylate
P:1-1938/Product: myosin heavy chain #status experimental <MAT>
P:89-768/Domain: myosin motor domain homology <MYOT>
P:179-186/Region: nucleotide-binding motif A (P-loop)
P:657-679/Region: actin binding #status predicted
P:841-1938/Domain: coiled coil <COI>
P:841-1289/Region: S2
P:852-1108/Domain: short subfragment 2 <SUB2>
P:1290-1938/Region: light meromyosin
P:1/Modified site: acetylated amino end (Ala) #status experimental
P:35/Modified site: N6-methyllysine (Lys) #status experimental
P:130,551/Modified site: N6,N6-trimethyllysine (Lys) #status experimental
P:185/Binding site: ATP (Lys) #status predicted
P:697,707/Active site: Cys #status predicted

F:755/Modified site: 3'-methylhistidine (His) #status experimental

Query Match 9.5%; Score 256; DB 1; Length 1938;
 Best Local Similarity 22.4%; Pred. No. 0.0015;
 Matches 121; Conservative 83; Mismatches 178; Indels 158; Gaps 18;

QY 33 SOAPKPEGAQRTAQSGALRVSELSRQLEDILSTVCVNNQGGPGEDGAQGPAPPE 92
 DB 1090 SOIQKIEDEALQVQ---LQKKIELQRIEEL-----EEIEAERTSRA 1132

QY 93 DAEKRTYVARNGEPEPTPVVYGEKPSKGDPTNTEIRQSDVGVDRHRRPOEKKAKGL 152
 DB 1133 KAEKTRADLSRELE-----ISERLEAGGATAAQI-----EMTKREAFQKMRD 1179

QY 153 GKEITILLMTLNTLSTPEKLAALCKY---AELESHENSOKMKLQKQSOLOVSK 208
 DB 1180 LEEATL-----QHEATAALRRKHADSTAEIGEQIDNIQRVKQKLEKSELMKEI 1230

QY 209 DILRGHSHKAVLARSKLESICREL-----QRNRSLSKESGVQVQAR----- 248
 DB 1231 DDLASNESVSQKAKANLEKWCITLEDQLSIEIKTEEQNQRMINDLNTQARLQETGEYS 1290

QY 249 -----BBEKKKEVTSH-FQVTLNDILOMEQH 276
 DB 1291 QRAEBKDALISQSRGKQFTQIEBELKXHLSEETKAKVALAHALQSAHDCDLLEQYE 1350

QY 277 ER-----NSKLR-----QENNELAERLKLIEQVELFEHEDKVF 311
 DB 1351 EQEAKGELQALSKANSEVQWRTKYETDAIQRTTELEAKKLAQRLQDAEHV----- 1406

QY 312 KHKDLOQVLDVAKLQAOQENLKEAERHOREKDFLKEAVESQRMCKMQQETHLKOQL 371
 DB 1407 -----EAVNAKAS-----LETKQRLQNEVEDLAVDVERSNAACAALDKKQKPFKIL 1455

QY 372 ALYTEKFEFQNTL-----SKSEVF---TTPKQ---EMEKMTKKIKLEKETTMYR 417
 DB 1456 AKWKQIBETQTELEASQKESLSLTFELFWKNAVEESLDHLETIKENKNIQOELADLT 1515

QY 418 SRWESNKALLEMAEKTVDKLEGLQVKIQRL-----KLCRALQTERNDLNKRV 469
 DB 1516 EQIAEGGKVAHELVKVKHVEQKSELSQALBEAASLSHERGKILR-LQLEINQIKSEI 1574

RESULT 11
 A38194
 X:Contains: desmoplakin II
 C:Species: Homo sapiens (man)
 C>Date: 04-Mar-1993 #sequence revision 03-May-1996 #text_change 21-Jul-2000
 C:Accession: A38194; A35536; B35536
 R:Virata, M.L.; Wagner, R.M.; Parry, D.A.; Green, K.J.
 A:Title: Molecular structure of the human desmoplakin I and II amino terminus.
 Proc. Natl. Acad. Sci. U.S.A. 89, 544-548, 1992
 A:Reference number: A38194; MUID:92115697; PMID:1731325
 A:Accession: A38194
 A:Molecule type: mRNA
 A:Residues: 1-974 <VR>
 A>Note: sequence extracted from NCBI backbone (NCBIN:75984, NCBI:75986)
 R:Green, K.J.; Parry, D.A.D.; Steinert, P.M.; Virata, M.L.A.; Wagner, R.M.; Angst, B.D.;
 J. Biol. Chem. 265, 2603-2612, 1990
 A:Title: Structure of the human desmoplakins. Implications for function in the desmosome
 A:Reference number: A35536; MUID:90153880; PMID:1689290
 A:Accession: A35536
 A:Molecule type: mRNA
 A:Residues: 'R', 927-2677 <GRB>
 A:CROSS-references: GB:J05211; NID:g181607; PIDN:AAA35766.1; PID:g181608
 A:Accession: B35536
 A:Molecule type: mRNA
 A:Residues: 'R', 927-1000, 1600-2677 <GR2>
 A:CROSS-references: GB:J05211
 C:Comment: Desmoplakins I and II, products of single gene, are major proteins of the int
 C:Genetics:
 A:Gene: GDB:DSP

A:CROSS-references: GDB:126564; OMIM:125647
 A:Map position: 6pter-6p21
 C:Keywords: alternative splicing; cytoskeleton
 F:1-2677/Product: desmoplakin I #status predicted <MAT1>
 F:1-1000,1600-2677/Product: desmoplakin II #status predicted <MAT2>

Query Match 9.5%; Score 256; DB 2; Length 2677;
 Best Local Similarity 23.2%; Pred. No. 0.0021;
 Matches 133; Conservative 107; Mismatches 195; Indels 138; Gaps 25;

QY 26 EAGPGSGSQAPRP-----EGQAATAQS-----GALRVSELSRQLEDIL 67
 DB 1238 QOKATGSEVSKQOQLEVLQVTFQMTESVRYKQSLDDAKTIQDKKSIER-LKOLI 1296

QY 68 STYCVNNQGGPGEDGAQGPAPPEDAEKST-VVARNGEPEPTPVV-----YGEKPSKG 122
 DB 1297 DKEINR-----KLEENARLQVQDLOKANSSATETINKLAVQEQELREL 1344

QY 123 DPNTTEIRQSDVGVDRHRRPOEKKKAKGLQKEITLLMTLNTLSTPEKLAALCKYAE 182
 DB 1345 RIDYERVVSQERTVADQDITRFQNSLKELOLQK--KVEEELNRLKRTASEDSCKRKLLE 1402

QY 183 LLESHENSQKMKL-----LQKKQSO--LVQEKD-----HLR-GEHSHKAVLAR 222
 DB 1403 ELEGRRSLKEQAIKITNLTOOLEQASIVKRSSEDDLRQDRDVLGHLREKQRTQSELRR 1462

QY 223 --SKLESICRELQHRNLSKE-----EGVQPARBEEKKEVTSFQVTLNDILOMEQH 275
 DB 1463 LSSEVEALRRQLQOESVQKQHLRNEHFQKAIEDKSR-----SLNESKTEIERL 1512

QY 276 NERNSKLQENKELAEKLL-IEQYELR-----BEHIDK-----VFHKDLQOQ 318
 DB 1513 QSLTENLTKHELMLEELNRLLEVDLRRGSRADSDKATILELSQLOISNNRTLEL 1572

QY 319 QLVDAKLAQAOQEMLKEAERHOREKDFLKEAVESQRMCK-ELMKQOETHLQKQALAYTEK 377
 DB 1573 QGLINDLQRENLRQETEFQKQKALEASNIQESKNQCTQVQVERESLLVKIKVLEQDK 1632

QY 378 -----PEFQNTLSKSEVFTFPQEMKMTKKIKK-LEKETTMYRSWESSNKALLEMA 431
 DB 1633 ARLOQLEDELNRAKSTLEAETRVKQRLCEKQKQIQNDLNQWKTQYSRKEAIRK--IESE 1690

QY 432 EEXTVRDK-----ELEGLQVKIQRIEKLICR-----ALQTERNDLNKRVQDLISA 474
 DB 1691 REKSEKNSLSRSEIERLQAEIKIEERCRRKLESDSTRETQSQLETSRYQREIDKL-- 1748

QY 475 GQGSGLTDSGPERPEPGCAQAPSPRVTEAPC 507
 DB 1749 -----RQRPYG-----SHRETQTEC 1763

RESULT 12
 T34021
 protein kinase SK2 - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34021
 R:Fukami, Y.; Yamamoto, H.; Ichihara, T.; Mori, K.; Gomi, T.; Sato, K.
 Submitted to the EMBL Data Library, April 1997
 A:Description: SK2, a putative rat homologue of yeast protein kinase NRK1.
 A:Reference number: Z21463
 A:Accession: T34021
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1206 <FK>
 A:CROSS-references: EMBL:AB003357; NID:d1106513; PID:d1020890; PIDN:BAA20077.1
 A:Experimental source: strain SD
 C:Genetics:
 A:Gene: SK2

Query Match 9.4%; Score 253; DB 2; Length 1206;
 Best Local Similarity 22.2%; Pred. No. 0.0013;
 Matches 133; Conservative 92; Mismatches 199; Indels 176; Gaps 25;

QY 12 EGQERPSQAAPVAEAGPSSQA-----PR 37
DB 682 ENAQELPVKAEP-----QAPAAQASEPPVLPISINHSNTENKMGALPKETILPP 737
QY 38 KPEGAQARTASG---ALRDVSELSQLELDILSTYCVDNNGGPGEDGAGQGEPAEPEDA 94
DB 738 EPENKGNDDTSGTSTVENSSDLNLISFSL-----KTKDSGSVSIQETRRQKTL 791
QY 95 EKRTYVARNGEPTPVVYGEKPSKGDNTTEIROSDVGDORDHRRPOEKKKAKGLSK 154
DB 792 KTKRFTVDGVEVSVT-----SKIVTSDSDKTELR-----FLRQELSELRLLOK 838
QY 155 EITLLMQTLN-TLSTPEKLAALCKYAELEHNSQKQMLLQKQSOVLQVQKDHRLG 213
DB 839 EEQAQOOLKGLQOQEQIFRFEQ--EMLSKKQYDQIENLEKQKQKTIK-----889
QY 214 EHSKAVLARSKLSCLRELQHNLSKBERGVORAREEERKKEVTSHPVTLNDIQLQME 273
DB 890 -----RLE-----QEHNLRLDE-AKRIKGEQKE-----LSKFNQMLN-----923
QY 274 QHNSKLNQENNELAKLLEQ-----YELREEH 306
DB 924 RXGEEQFVQKQOQELQALKKIQQKQKALANIERECLNKKQOLLREARAAIWELEERH 993
QY 307 IDKVFQKHDLQOQVLDAKLAQAQEMLEAE---ERHQEKDFLLKEAV--ESQSMCELMK 361
DB 984 LQE--KHQLAQQLKQVFIQRHQLLKHEKETEMQKYNQRLTEELKNRQTOBRARLPK 1041
QY 362 QOETHLQOALY-----TEKFEFQNTLSKSEVFTTFKQEMKTKKI 406
DB 1042 IORSEAKTRAMFKKSLINSTATPDQDREKIKQFAAQEEK-----ROKNRMAQH- 1092
QY 407 KKLEKTTMYSRVSSSKALLEMAEEK-----TVRDKELR---GLQVKIQRLEKLC 455
DB 1093 QKHESQMRDLQLOCEANVRELHQLQNEKCHLVEHETQKLKELSDHESQELKEWR-SKLR 1151
QY 456 RALQTERNDLNKRVODLS-----AGGQSLSDSGPERRPQGAQSSRVTEAPCYFG 510
DB 1152 PRKKTLEEFARKLQEQEVFFMTGSECLNPSAQSR-----GCLQTSHPSSSTAPAWAG 1206
RESULT 13
P75103
conserved hypothetical protein PAB0812 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 02-Feb-2001
C:Accession: F75103
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: F75103
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-880 <KAW>
A:Cross-references: GB:AJ748286; GB:AL096836; NID:9545836; PIDN:CAB50131.1; PID:9545864
A:Experimental source: strain Orsay
C:Genetics:
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032
Query Match 9.3%; Score 252.5; DB 2; Length 880;
Best Local Similarity 23.2%; Pred. No. 0.00095;
Matches 123; Conservative 98; Mismatches 183; Indels 127; Gaps 22;
QY 12 EGQERPSQAAPVAEAGPSSQ-----APRKPEGQAARTAQGALRDVSELSRQLE 64
DB 259 KGLEEKIVQIERSIEBKAKISELEIVNDIPKLOEKEKRYKLKG-FRDEYESKLRLLE 317
QY 65 DILSTYCVDNNGGPGEDGAGQGEPAEPDAKSRITYAENGPEPTPVVYGEKPSKGD 124
DB 318 KELSKEW-----ESELNAIEVTKGB-----KKKE 342

QY 125 NTEIROSDVGDRTDHRP--QEKKAQGLGKEITLIMQTLNTLSTPE--EKLAALCK 178
DB 343 RABEIREKLSIEKLEHLKPYVELEDAKQVQKIERKARLKGSLSPGEVIEKLESLEK 402
QY 179 KYAELEHNSQKQMLLQKQSOVLQVQKDHRLQREHSHKAVLARSKLSLC-----REL-Q 233
DB 403 ERTEI-----EEAKIETTRIGQMEQEK-----ERMKAIBELKAKGKCPVQRELU 451
QY 234 RHNRSIK-----BEGVORAREEERKKEVTSHPVTLNDIQLQ-----MEQHNRNS 280
DB 452 EHKELMERYTLEIKIIEELARTTEERKLVNLRLEIKLREPSVMDIAEQIKELSS 511
QY 281 KLROENMELAE-----RLKGLIEQYELREHDKVPK 312
DB 512 KLKGFNLELEQKEREFEGLNEEFNKLKGEILGLERDLAKRIKALEGRRKLEEKVKRKAKE 571
QY 313 K-KDLOOQLVDKLAQAQEM---LKEAERHOREKDFLLKEAVESQRMCELAQOETHLK 368
DB 572 ELNHRQRLGFGFSEBELNRIQLESEFHDYKVEAKKSES-ELRELKKNLEKKEKTELD 630
QY 369 QQLALYTERKFEFQNTLSKSEVFTTF-KOENKMTKKIKLEKETMYTRSWESSNKA- 426
DB 631 QAFENLADVENIEEKEAKLDESKEFNEBEYERKRLVKLEREVSSITARLEBLKSV 690
QY 427 -----LLEWAERKTVRDKLEGLQVKIQRLEKLCALQTERNDLNKRVOD 471
DB 691 EQIKATLRKLKSEKEREK-----AKLEIKKLEKALSKEV-----DLRKKIND 733
RESULT 14
A40691
trichohyalin - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C>Date: 21-Sep-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C:Accession: A40691; A34209; S32633
R:Pietz, M.J.; McLaughlan, C.J.; Campbell, M.T.; Rogers, G.E.
J. Cell Biol. 121, 855-865, 1993
A:Title: Analysis of the sheep trichohyalin gene: potential structural and calcium-bind
A:Reference number: A40691; MUID:93260018; PMID:7684041
A:Accession: A40691
A:Molecule type: DNA
A:Residues: 1-1549 <PIE>
A:Cross-references: EMBL:Z18361; NID:g295940; PIDN:CAA79165.1; PID:g295941
A:Note: sequence extracted from NCBI backbone (NCBIP:132511)
R:Pietz, M.J.; Presland, R.B.; Rogers, G.E.
J. Cell Biol. 110, 427-436, 1990
A:Title: The cDNA-deduced amino acid sequence for trichohyalin, a differentiation marke
A:Reference number: A34209; MUID:96130632; PMID:2298812
A:Accession: A34209
A:Molecule type: mRNA
A:Residues: 1016-1151, 1205-1257, 1281-1398, 'G', 1400-1549 <PI2>
A:Cross-references: GB:X51695; NID:g1827; PIDN:CAA35992.1; PID:g1828
C:Comment: trichohyalin is a protein of the medulla of the hair and of the inner root s
Covalent modifications to this protein include conversion of arginine to citrulline and
C:Genetics:
A:Introns: 46/3
A:Note: single copy gene
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
P:49-81/Domain: calmodulin repeat homology <EP2>
P:387-851/Region: 28-residue repeats
P:886-1519/Region: 23-residue repeats
Query Match 9.3%; Score 251.5; DB 1; Length 1549;
Best Local Similarity 22.7%; Pred. No. 0.0019;
Matches 111; Conservative 108; Mismatches 194; Indels 77; Gaps 17;
QY 6 QPEAGPQAQRPQAAPVAEAGPSSQAPRKPEGQAARTAQGALRDVSELSRQLE 65
DB 521 QREERKROERQYLEKVELOEEQLOQREKREKQKQ-----YLEKVELQEE 574
QY 66 ILSTYCVDNNGGPGEDGAGQGEPAEPDAKSRITYAENGPEPTPVVYGEKPSKGD 125

QY 374 YTEK-----FBEFQNTLSSSEVFTTFKQEMEMTKKIKKLEKETTMYRSRWESSNKAL 427
Db 1483 FPAKQSDADLELNKTVSEHDBVAKLNTQITFLTRDNQSAEELNELRSKADKDKKKI 1542
QY 428 LEMAE-----EKTVD-----KLEGLQVKIQRLEKLCRALQTER 462
Db 1543 SELEEQVNELESFVGTGNADENEIRKDAQIADLNKALEMGVQVQVQVQVQVQVQVQVQV 1602
QY 463 NDNLKRVQ 470
Db 1603 NDLTSKIE 1610

RESULT 17

A43336
microtubule-vesicle linker Clip-170 - human
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-Nov-1999
C:Accession: A43336
R: Pierre, P.; Scheel, J.; Rickard, J.E.; Kreis, T.E.
Cell 70, 887-900, 1992
A:Title: CLIP-170 links endocytic vesicles to microtubules.
A:Reference number: A43336; MUID:92405160; PMID:1356075
A:Accession: A43336
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1392 <PIR>
A:Cross-references: GB:M97501; NID:g180621; PIDN:AAA35693.1; PID:g180622

Query Match 9.2%; Score 249; DB 2; Length 1392;
Best Local Similarity 24.6%; Pred. No. 0.0021;
Matches 115; Conservative 83; Mismatches 152; Indels 118; Gaps 20;
QY 117 KEPSKGPDPNTEIRQSDVDR-----DHRPQKKKAGLGE-----ITLLMQTLN 164
Db 735 KASSEKSEMKKLRQQLQAEAKQIKHLIEKNAESSKASSITRELOGRRLKTLNQLNLS 794
QY 165 TLS-----TPBEKLAALCKKYAEILEE-----HRNSQOMKLLQKQSQOL----- 204
Db 795 EVSOVKETLEKELQILKEKFAEASEAVSVQSRQETVVKLHQKBEQFNLSSDLKLE 854
QY 205 -----VQKHRLGEHSAVKLARSKLSLCRELQ-----HNR-----LKPEG 243
Db 855 NLADMEAKPREKDE-----REEQLKAKEKLENDIAEIMXSGDNSSOLTKNDLRLKERD 911
QY 244 VQRAEBEEKEKVTSHFQVTLNDIOLQMEQ-HNERNKLRQENMELARLKLLEQYEL 302
Db 912 VBEQLKLTVANASFLQKSIDMTVRAEQSQEAAKKEBEKELERKLSLEKKMET 971
QY 303 REEHIDKV-----FKHKO-----LQQLQVDA--KLQQAQE-----MLKEAEE-RH 339
Db 972 SHNQCELKARYERATSETTKHEEILQNLQKTLTDKLGARENSGLLQLEELRK 1031
QY 340 QREKDFLLKEAVESQRMCELMKQ-----ETHLQQLALYTEKFEFQNT--L 385
Db 1032 QAEAKAAQTAEDAMQVQMTKTEKTLASLEDTKQTNKALQNELDTLKNLKNVEEL 1091
QY 386 SKSEVFTTFKQEMEMTKKIKKLEKETT-----MYRSRWESSNK 426
Db 1092 NKSLEITVENQKMEBEFKSEITLKQAAQKSQLSALQENVKLAELGERSDEVTSHQ 1151
QY 427 LLEMAEKTVKDEKLEGLQVKIQRLEKLCRALQTERDIAKRVQDLSA 474
Db 1152 KLE--EERSVLNQL--LEMK-KRESKFTKADBEKASLQKTSITSA 1194

RESULT 18

A59293
skeletal myosin heavy chain - domestic rabbit
C:Species: Oryctolagus cuniculus
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000
C:Accession: A59293

R:Maeda, K.; Hostinova, B.; Roesc.Kleinkauf, A.; Schuster, H.; Gasperik, J.; Wittinghof,
submitted to GenBank, July 1995
A:Description: Isolation, sequencing of myosin heavy chain cDNA from rabbit skeletal mu
A:Reference number: A59293
A:Accession: A59293
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1938 <MAE>
A:Cross-references: GB:U32574; NID:g940232; PIDN:AAA74199.1; PID:g940233
A:Experimental source: strain New Zealand White; cell type skeletal muscle fiber type I.
C:Genetics:
A:Gene: MYH
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:89-769/Domain: myosin motor domain homology <MMO>

Query Match 9.2%; Score 249; DB 2; Length 1938;

Best Local Similarity 21.2%; Pred. No. 0.003;
Matches 113; Conservative 80; Mismatches 179; Indels 160; Gaps 16;
QY 33 SQAPKPGCAQARTAQSGALRDVSELSRQLEDILSTCYVDNNQGGPGDGAQGSFAPRE 92
Db 1091 SNLQSKIEDQALAMQ---LQKKIKELQARIEEL-----EEIEAERASA 1133
QY 93 DAEKSTYVANGPEPEPTFVYVYGEKPSKGPDPNTEIRQS-DEVDGRHRRPOEKKKAG 151
Db 1134 KAEKQSDLSR-----ELEISERLEEAGGATSAQIEMMKRBA 1172
QY 152 LKKEITLLMQTLNLTSTPEEKLAALCKK-----VAILLEHRNSQOMKLLQKQSQOLVQ 206
Db 1173 ---EFQKVRDLLEAATLQHEATNATLRKHADSVAILGEQIDNLRQVKLEKESLKM 1229
QY 207 EKDHRLGEHSAVKLARSKLSLCREL-----QRHNRSLKEGVQAR----- 248
Db 1230 EIDDLASNNETVSKAKGNLKMCRITLEDQVSELKTEBEHQRLINDLSAQRARLQTESG 1289
QY 249 -----EEREKKEVTSH-FQVTLNDIOLQMEQ 274
Db 1290 FSRQLEKDSLVQSQRGQAFTQOIEELKQLESEIKAKSALALQASRDCDLRLQ 1349
QY 275 HNER-----NSKLR-----QENMELARLKLLEQYELREHIDK 309
Db 1350 YEEQEAKAEQRAKMSKANSVAQWRKYETDAIQRTBELEAKKKLQALQDAEHHV-- 1407
QY 310 VFKHDKQQLQVDAKLQQAQEMELKASERHQREKDFLLKEAVESQRMCELMKQOETHLKQ 369
Db 1408 -----EAVNAKAS-----LEKTKQRLQNEVEDIMDVERTNAACALDKQGNFK 1454
QY 370 QLALYTEKPEFQNTLSKSSSEVFTTFKQEMEMTKKIKKLEKETTMYRSRWESSNKAL 429
Db 1455 ILAEKHKVETHT-----AELEASQKESRSLSTEVEFKVKNAYEESLDQLET 1500
QY 430 MAEKTVDKLEGL--QV-----KIQRLKLCRALQTERDIAKRVQDLSA 474
Db 1501 LKRENKQLQEEISDLTEQIABGGKRIHLEKVKQVQBEKSELQALBEA 1552

RESULT 19

T17365
serine/threonine protein kinase TA01 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17365
R:Hutchison, M.; Berman, K.S.; Cobb, M.H.
J. Biol. Chem. 273, 28625-28632, 1998
A:Title: Isolation of TA01, a protein kinase that activates MEKs in stress-activated pr
A:Reference number: Z18730; MUID:99003202; PMID:9786855
A:Accession: T17365
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1001 <HUT>
A:Cross-references: EMBL:AF084205; NID:g3452472; PID:g3452473; PIDN:AAC71014.1
A:Description: probably implicated in the regulation of the p38-containing stress-respo

```
Query Match          9.1%; Score 247; DB 2; Length 1001;
Best Local Similarity 21.9%; Pred. No. 0.0018;
Matches 142; Conservative 118; Mismatches 212; Indels 176; Gaps 30;

QY 23 PAVEAB-----GPSSQA-PRKPGAQARTAQSGALRDVSELSROLE 64
DB 324 PAVEAQEEBEOHGGRTGTNSVGSNOSIFSMSISASSQSSVNSLFDASDDKS-ELD 382
QY 65 DILSTYCYVNNQO---GFCBGAQGEPAEPDAEKSRTYTVARNGEPEPTPVVYGEKSPS 120
DB 383 MNEGDHTVNSNSVIHLKPEENYO-EGDPR-----TRASAPQPPQVSRHKSXY 432
QY 121 KGPNTBETROSDEVDRHRRPOEKKAKGLGKEIT-----LLMOTLNTLSTPEEKALAL 176
DB 433 RNREHPATITASLV---TRQCHQEHSLEKQSGYKRRROKQKQMLTENKIKAB 488
QY 177 CKKYA-----ELLFEHFN--SQKQKMLLOKQKQSOLOVE-----KDHLRGERSKAV 219
DB 489 MDEHRLRLDKDLTQNNPFAEMEKLIKQASMEKEAKVMANERKKPQCHIQAQOKKEL 548
QY 220 LARSKLESICRELQRNRSILKEGVOR-----AREEE---KRKEVTSHEQV--TLNDIQL 270
DB 549 --NSFLESQKREYKLRKEQLKXELNENQSTPKKEQEWLSKQKENTOHFOABEENILRR 606
QY 271 QME-----OHN-----ERNSKLRQENMELAEKLIKLIQVYELREEH 306
DB 607 QROYLELCRPRKRMELGRHNEQDLVRELNKRTQKLEHMLLRQHESQLELFRH 666
QY 307 IDKVPKHDLOQOLVDKLOQAQOEMLKEABERHOREKDFLLKEAVSORMCELMKQOETH 366
DB 667 LNTIQK---MRCBLL--RLQHOTELTNQLEYNRRERELRKEVMEVROQPKSLKSKELQ 721
QY 367 LKQOL---ALYTEKEEFONTLSKSEVFTTFOQEMKTKIKKLEKETTMYRSRWES 422
DB 722 IKQFOFTCKIQROYKALRNHLLE---TTPKSEKAVLKXK--BEQTRKLAILEQ 774
QY 423 SNKALLMAEKEKTVR-----DKELEGL 444
DB 775 YDHSINEMLSQALRLDEBAEQCVLKMQLQOEBELNAYQSKIKQMAQABQDRELREL 834
QY 445 QVKI-----QLEKLCRALQTERND-----LNKEVQDLSAGGGS-----LTD 482
DB 835 EQRVSLRALLQEKIEEMALQOATERIRSLRQARIEAPDSESMLGFSNNVLSN 894
QY 483 SGPERRPEG-PGAQPS-SRVTBAP--CYPGAPSTBAG---QTGPQ 523
DB 895 LSPEAFSHSYPGASSWSNPTGSGPHGHPMGCTPOAGHPMGQGPQ 942

RESULT 20
S22695
restin - human
C:Species: Homo sapiens (man)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 05-Nov-1999
C:Accession: S22695; S19853
R:Biibe, G.; Delabie, J.; Brueggens, J.; Richener, H.; Asselbergs, F.A.M.; Cerletti, N.;
EMBO J. 11, 2103-2113, 1992
A:Title: Restin: a novel intermediate filament-associated protein highly expressed in th
A:Reference number: S22695; MUID:92289675; PMID:1600942
A:Accession: S22695
A:Molecule type: mRNA
A:Residues: 1-1427 <BL>
A:Cross-references: EMBL:X64838; NID:935998; PIDN:CAA46050.1; PID:935999
C:Keywords: cytoskeleton

Query Match          9.1%; Score 246; DB 2; Length 1427;
Best Local Similarity 24.4%; Pred. No. 0.0029;
Matches 114; Conservative 84; Mismatches 150; Indels 118; Gaps 20;

QY 117 KEPSKDPNTBEIROSDEVR-----DHRRPQKKYAKGLGKE-----ITLLMTQIN 164
DB 770 KASSBGSKMKLRQQLAEKQKHLKLEIKQAVSSKASSITRELQGLKLTNLQENLS 829
```

```
QY 165 TLS-----TPEKLAALCKKVAELLEER-----HRNSOKOMKLLQKKOSQL----- 204
DB 830 EVSQVKETLEKLOILKEKFAEAEAVSVORSQETVYKHLQKEEQFNMLSSDLKLE 889
QY 205 -----VQEKHLRGEHSAVLARSKLESCLSELQOR-----HNRS-----LKEEG 243
DB 890 NLADWEAKFREKDE---REEQLIKAKRLENDIAEIMQSGDSSQLTKMDELKQERD 946
QY 244 VQRAEEEEKEKEVTSHPQVTLNDIQLQMEQ-HVERNSKLQENMELAEKLIKLEQVEL 302
DB 947 VEELQKLTKANENASFLQKSIDMTVFAESQOQRAAKKHEEKLEKPLSDLEKKMET 1006
QY 303 REEHDKV-----FKHKD-----LQOQLVDA--KLQQAQOE-----MKAEAE-RH 339
DB 1007 SHNQOQLKARYERATSETTKGHEILLQKTLTDKLGKAEENSGLIQELEELRK 1066
QY 340 QREKDFLKEAVSORMCELMKQO-----ETHLQOOLALYTEKPEEPONT--L 385
DB 1067 QADKAKAAQTAADAMQIMEQTKETETLASLEDTKQTNALQNELDTLKENNLKXVEL 1126
QY 386 SKSSSVFTTFOQEMKTKIKKLEKETT-----MYRSRMSSSKNA 426
DB 1127 NKSKELLTVENQKMBEPRKEIETLKQAAQKSQLSALQAEENVKLAEBLGRSRDEVTSHQ 1186
QY 427 LLEMAEKTVRDKLEGLQVQKIQRLKLCRALQTERNDLNKRVQDLSA 474
DB 1187 KLE--EERSVLNNQL--LEMK-KRESKFIKDADESKASLOKSISITSA 1229

RESULT 21
I38055
myosin heavy chain, perinatal skeletal muscle - human
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Homo sapiens (man)
C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 19-Apr-2002
C:Accession: I38055; JH0154; S12459; S09332; A30220; S49478
R:Bullian, E.H.; Kelly, A.M.; Pompidou, A.J.; Hoffman, R.; Schiaffino, S.; Stedman, H.H.
Eur. J. Biochem. 230, 1001-1006, 1995
A:Title: Characterization of a human perinatal myosin heavy-chain transcript.
A:Reference number: I38055; MUID:95324556; PMID:7601129
A:Accession: I38055
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1937 <RS>
A:Cross-references: EMBL:Z38133; NID:9558668; PIDN:CAA86293.1; PID:9558669
R:Karsch-Mizrachi, I.; Peghali, R.; Shows, T.B.; Leinwand, L.A.
Gene 89, 289-294, 1990
A:Title: Generation of a full-length human perinatal myosin heavy-chain-encoding cDNA.
A:Reference number: JH0154; MUID:90323631; PMID:2373371
A:Accession: JH0154
A:Molecule type: mRNA
A:Residues: 1-14, 'A', 16-859 <KAR>
A:Cross-references: GB:Y00821
A:Experimental source: skeletal muscle
R:Bober, E.
submitted to the EMBL Data Library, January 1989
A:Reference number: S12458
A:Accession: S12459
A:Molecule type: mRNA
A:Residues: 502-1071, 'N', 1073-1250, 'DGG', 1253-1376, 'NT', 1379-1913, 'D', 1915-1937 <BOB>
A:Cross-references: EMBL:X51592; NID:929465; PIDN:CAA35941.1; PID:929466
R:Bober, E.; Buchberger-Seidl, A.; Braun, T.; Singh, S.; Goedde, H.W.; Arnold, H.H.
Eur. J. Biochem. 189, 55-65, 1990
A:Title: Identification of three developmentally controlled isoforms of human myosin he.
A:Reference number: S09331; MUID:90235862; PMID:1691980
A:Accession: S09332
A:Molecule type: mRNA
A:Residues: 502-547, 'X', 549-617, 'X', 619-687, 'X', 689-757, 'X', 759-827, 'X', 829-897, 'X', 899-1376, 'NT', 1379-1386, 'X', 1388-1456, 'X', 1458-1526, 'X', 1528-1596, 'X', 1598-1666, 'X', 1668-1
A:Cross-references: EMBL:X51592
R:Peghali, R.; Leinwand, L.A.
```


C:Species: Mus musculus (house mouse)
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999
 C:Accession: S49461, S59599
 R:Julien, S.; Luc, M.; Francois, C.
 Submitted to the EMBL Data Library, October 1994
 A:Description: Cloning and sequencing of the murine SCP1 cDNA.
 A:Reference number: S49461
 A:Accession: S49461
 A>Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-993 <JUL>
 A:Cross-references: EMBL:238118; NID:g1360015; PIDN:CAA86262.1; PID:g558603
 Biochim. Biophys. Acta 1263, 258-260, 1995
 A:Title: cDNA sequence of the murine synaptonemal complex protein 1 (SCP1).
 A:Reference number: S59599; MUID:96004899; PMID:7548215
 A:Accession: S59599
 A>Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-993 <SAG>
 A:Cross-references: EMBL:238118; NID:g1360015; PIDN:CAA86262.1; PID:g558603

Query Match 9.0%; Score 243.5; DB 2; Length 993;
 Best Local Similarity 22.9%; Pred. No. 0.0026;
 Matches 124; Conservative 91; Mismatches 183; Indels 143; Gaps 23;

QY 34 QAPRPEGAQARTAGSALRDVSELSKQLSD-----ILSTVCVD 73
 DB 132 QENRRIIAQRAIOE--LOFENEKVSLSKEEIQENKDLIKENNATHTWNLKETC-- 187
 QY 74 NNQGGGEGDGAQGEPAEDAKSTYVYARNGEPEPTVVYGEKPSKGDPTNEEIRQSD 133
 DB 188 -----ARSAETKNKYVERETRGVYVDLSNIEKILAF-----EELVQA 229
 QY 134 EVG-----DRDHRFPQ--BKQAKGLGKEITLLMOTLNTLSTPPEKLAALCKYVBL 183
 DB 230 ENARLEMFKLKEDHEKIQHLEEEYQKEVNNKENQVSELLTQSAEKENKM-----KOLTF 285
 QY 184 LEEHNSQKQM-----KLQKQSQVQKQHLRGHSHKAVLARSKLSLCEELQRH---N 236
 DB 286 LEESDKANLEKTKLQDENLKESEKQHL-----TSELDIKMSHORSSTQ 335
 QY 237 RSLKEE-----GVQAREEERKEVTSHPQV-----TLNDIQLOMBQHN 276
 DB 336 KALEEDLOIATNTISQLETVKAEQWELNKAHTHSFVVTLEKATCTTLELLRTQOOL 395
 QY 277 ENNS-----KLQENNELAERLKKLIQVYELREEHIDKVPKHDLQOQLVDAKLOQA 328
 DB 396 EKNEPQLKLTVELQKKNEL--EMTKFNKNEVELBELKNILAE--DQKLDEK-KQV 450
 QY 329 QMLKAEERHQREKDFLL-----KEAVESQWCELMKQOETH-LKQQLALYTE-KFEEFPQ 382
 DB 451 EKLAELQKEQ-ELTFLLETRKEVHDQEQVTVTKTSQHYLKQVEEMKTELEKEKLX 509
 QY 383 NT-----LSKSSSEVPTTFKQEMKMTKKIKLEKETMY 416
 DB 510 NTELTASCDMLLENKKTFQASDMALEKXQHQDIINCKQGEERLQKLENELEKMH 569
 QY 417 RSRWESSNKALLEMAEE-KTVDRKLE---GLQVKIQLEKLCALQTERNDLNKRVQDL 472
 DB 570 RDELESVRKEFTQOQDEVKCKLDKSEENARSIECEVLKCKEKQKILSKCNLKKQVENV 629
 QY 473 S 473
 DB 630 S 630

RESULT 24
 A56734
 ribosome receptor, 180k - dog
 C:Species: Canis lupus familiaris (dog)
 C>Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 08-Oct-1999
 C:Accession: A56734

R:Wanker, E.B.; Sun, Y.; Savitz, A.J.; Meyer, D.I.
 J. Cell Biol. 130, 29-39, 1995
 A:Title: Functional characterization of the 180-kD ribosome receptor in vivo.
 A:Reference number: A56734; MUID:95310363; PMID:7790375
 A:Accession: A56734
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1534 <WAN>
 A:Cross-references: GB:X87224; NID:g984113; PIDN:CAA60676.1; PID:g984114
 C:Keywords: endoplasmic reticulum; membrane protein; protein biosynthesis
 F:198-743/Region: 10-residue repeats (N-Q-G-K-K-E-G-A-P)

Query Match 9.0%; Score 242; DB 2; Length 1534;
 Best Local Similarity 20.4%; Pred. No. 0.0046;
 Matches 153; Conservative 114; Mismatches 222; Indels 260; Gaps 30;

QY 1 KSSPGQPEAGPR-----GAQERPSQAPAVEAEGPGSSQAP 36
 DB 710 KKSEGPSNQGKVDASANQSKRAESAPITGGKNADVQSQEAPKQZAPKCKSGSKKGE 769
 QY 37 RXP-----EQAQAR-----TAQSGALRDVSEE----- 58
 DB 770 GPPDSPLPYLKYTLVSTVGVNFEQEAQRLIILLSEKAGVTDTHKATQKQDPVAL 829
 QY 59 LSRQLED-----ILSTVCVNNQGGGEGDGAQGEPAEDAKSRT---YVARGEPETFPV 112
 DB 830 LKROLEEKELIAT-----EQEDAAVAKSKLREVNKELAAEKAKA 869
 QY 113 VTGEPEPSK---GDPNTEETIRQSDVGDGRDHRPQEKKKAK-----G 151
 DB 870 AAGEAKVKQLVARQBITAVQARIRASVREHVQVQLOQKIRTLQOLENGENTQLAR 929
 QY 152 LKQETILLMOTLN-----TLSTPEKLAALCKYAEELLEHRNSQKQMLQK 199
 DB 930 LOQENSIILDALNQATSVQESKQNTLAKLQKSELVSEKSEAAARQEQOKALLET 989
 QY 200 KQSQVQKQHLRGHSHKAVLA-RSKLSLCEEL---QRHNSLKEEGVQVAREEERKX 255
 DB 990 KTAALKEQVLQASHKESSEALQRLDEVSRELCSRTSHASLRAD-AEKARQOQOQMA 1048
 QY 256 EYTSHFQVTLNDIQLOMBQHNERSKL---RQBNWELAEELKK---LIEQYELRE----- 304
 DB 1049 ELHSKLQSSRAEVKSKSELSGLHGLKAEARNSQLMRIRSTEALLEGQARDTQDAQ 1108
 QY 305 -----EHIDKVPKHDLQOQLVDAKLOQAQCEMLKBAEERHQ-----REKDFLLKEAVES- 353
 DB 1109 ASRAEERQAL---KELESQV--WCLEKBATLKEAVQKVKVNDLREKNWKAMEALASA 1163
 QY 354 QEMCELMKQOETHLKQO-----LAL-----YTE----- 376
 DB 1164 ERACHEKLSRUTQAKBSEKOLSLTRACTKEALLALLPALSSAPQSYTEMLOELREKGP 1223
 QY 377 -----KFEFQNTLSKSSEVTTFKQEMKMTKKIKK-LEKETT 414
 DB 1224 ELKQRPADTPSSDLASKLREAEETQNNLQACDQVRTILAEETGLMKOLQKSVBEEQ 1283
 QY 415 MYRSWESSNKALLEMAEEKTVDRKLEGLQVKIQ-----BLEKLCR 456
 DB 1284 V-----WKAKVATTEELKQSKRVTVYKHLEDIVKELGELSESSEQVREHTSHLEAELEKHA 1339
 QY 457 ALQTERNDLNKRVQDLGAGGQCSLTDSGPERRRPGQAGAPSS-----PRVTEAPCY-- 508
 DB 1340 AASAEQCSYAKEV---AGLEQLLLESQSLDAKSEAKQSQNELALVRQQLSEWKSHEV 1395
 QY 509 -----PGAPSTASGQTGPQEPSTARA 530
 DB 1396 DDVAGSPAAPPAB-----QDPVELKA 1417

RESULT 25
 A56539
 giantin - human
 N:Alternate names: macrogolgin

C:Species: Homo sapiens (man)
 C>Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999
 C:Accession: A59234
 R:Seelig, H.P.; Schratz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
 Mol. Cell. Biol. 14, 2564-2576, 1994
 A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein
 A:Reference number: A59234; MUID:94187728; PMID:7511208
 A:Accession: A59234
 A:Molecule type: mRNA
 A:Residues: 1-3259 <SEE>
 A:Cross-references: EMBL:X75304; NID:9405714; PIDN:CAA33052.1; PID:9405715
 C:Genetics:
 A:Gene: GDB:GOLGB1; GCP: GCP371
 A:Cross-references: GDB:454958
 A:Map position: 3q13.31-3q13.31
 C:Superfamily: giantin
 C:Keywords: coiled coil; Golgi apparatus; transmembrane protein
 P:3238-3254/Domain: transmembrane #status predicted <TN>

Query Match 9.0%; Score 242; DB 1; Length 3259;
 Best Local Similarity 20.4%; Pred. No. 0.01;
 Matches 121; Conservative 99; Mismatches 169; Indels 206; Gaps 22;

QY 43 QARTAGALRVSELSRQLRDLITTC-----VNNQGGPGEDGAQGEPAEPEDAEX 96
 Db 1743 QSLSEKOSLSREVQDLKHQIEDVNSKQANLEATEKHQINVTIEGTQSIPTGETRQDS 1802
 QY 97 -SRTYVARGPEPT-----PVYGEKPPSGKDPNTEIROSDEVDGRDHRPPOEKKAK 150
 Db 1803 LSNSTRPTCSSEVPKSAKSNPAV--SKDPSSHDEINNYLOIDOLKRIAGLEBEKQNK 1860
 QY 151 GAGKEI-----TLAQ-----TLNLTS-----TPES-- 171
 Db 1861 EFSQTLNEKNTLLSQISTKQGLKWLQAEVTKWLLNQIQELBSRVTKLATAEEDK 1920
 QY 172 -----KLAL-----CKYA-----ELLE-EHNSQKQKMLQKQSQOLVQEKD 209
 Db 1921 DLERRLMNQLBELSGNIGYQDVDTAQIKNELLESKMLKCVSLEBEKQQLVKEKT 1980
 QY 210 HLRGHSHKAVLAR-----SKLSLCLRLQHRNSLK----- 240
 Db 1981 KVESEIRKEYLEKIQGAQKEPGNKSHAKELQELLKKEQEVKQKQDCIRVQEKISALER 2040
 QY 241 -----EQQVQARBEERKEKVTSHFQVTLNDIQLQWQHNSK 281
 Db 2041 TVKALFFVQTESQKDLTKENLAQAVHRKKAQALASFKVLLDDTQSEAAVLDNLK 2100
 QY 282 LRQENNELARLKLIEOVLELREHIDKVFKHQDLQQLVDKQLQAQELKEAEERHOR 341
 Db 2101 LKKELQSNKESVKSQWKQ-----KXEDLERR-----LEQAREKH-- 2134
 QY 342 EKDFLKEAVESQRCMLKQOETHLKQOLALYTEKEEFQNTLSKSEVFTTFKQEMEX 401
 Db 2135 -----LKEKNQWQELDALREKVLHETIG-----EIQVTLNKKDKVQQLQENLDS 2182
 QY 402 -----MTKKIKLEKETTMV---RSRWSSNKALLEMAEKEKTVRDKK----- 440
 Db 2183 TWTQALAAFTKMSLSLQDDRVDIDEAKWERKFSDAQSKSEE-IRLKEDNCSVLKQDLR 2241
 QY 441 -----LEGQVQKQRL-----KLCLALQTERNDLNKRVQD 471
 Db 2242 QMSIHHEELKINISLREHDKQIWESKAQTEVLOQKVCDTLQGENKELLQOLE 2295

RESULT 26
 A59234
 slow myosin heavy chain 3 - quail
 C:Species: Coturnix coturnix
 C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 08-Sep-2000
 C:Accession: A59234
 R:Nikovits Jr., W.; Wang, G.F.; Feldman, J.L.; Miller, J.B.; Wade, R.; Nelson, L.; Stock
 J. Biol. Chem. 271, 17047-17056, 1996
 A:Title: Isolation and characterization of an avian slow myosin heavy chain gene expressed

A:Reference number: A59234; MUID:96291845; PMID:8663323
 A:Accession: A59234
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1931 <NIK>
 A:Cross-references: GDB:U53862; NID:91289513; PIDN:AA59912.1; PID:91289514
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 P:81-761/Domain: myosin motor domain homology <MMO>

Query Match 8.9%; Score 240; DB 2; Length 1931;
 Best Local Similarity 21.7%; Pred. No. 0.0071;
 Matches 110; Conservative 92; Mismatches 182; Indels 122; Gaps 17;

QY 40 EGAQARTAGALRVSELSRQLRDLITTCVDDNQQGGPGEDGAQGEPAEPEDAERSKT 99
 Db 1090 EDEQALAAQ---LQKKLQELQARTEEL-----EELEAERTGRAKVEKLS 1132
 QY 100 YVARGPEPTPVYGEKPPSGKDPNTEIROSDEVDGRDHRPPOEKKKAGLKEITLL 159
 Db 1133 ELQ---ELET-----SERLEAGGATSVQL-----ELNKKREAEFQKLARDLEBATL- 1178
 QY 160 MOTLNTLSTPHEKLAALCKY---ASLEEHRSQKMLKQKQSQOLVQEKDHLRGEH 215
 Db 1179 -----QHEATAALRKCHADSVASLESEQLDNQVRVKQKLEKSEKLELDDVSSNM 1230
 QY 216 SKAVLARSKLSELCR-----ELQRHRSLSK----- 241
 Db 1231 EQLIKAKANLENKWCSTEDQWHERNKLESSQRTVTLSTORAKLOTENSLSRQLEBE 1290
 QY 242 -----EGVQARBEERKEKVTSH-FQVTLNDIQLQWQHNSKLR 283
 Db 1291 AFINQMRGKLTYYQQLDLARQLESEAKAKNALAHALQSAHQDCDLREQYBEE---M 1346
 QY 284 QENMELARLKL---IQVLELREHIDKVPKHQDLQ---QLVDAKLQQAQK----- 330
 Db 1347 EKAKLQALSKANSEVAQWTKYE-TDAIQTSELSEAKKLAQRLQAEAEVAVNAK 1405
 QY 331 --MLKEABERHQEKDFLLKEAVESQRCMLKQOETHLKQOLALYTEKEEFQNTLSK 388
 Db 1406 CSSLEKTHRLQNETLMDADVERSNAAAALDKQKFNFKILSEWQKQTEESQTELEAS 1465
 QY 389 SVFTTFPQKEKEMTKIKLEKETTMVRSWSSNKALLEMAEKEKTVRDKLEGLQVKI 448
 Db 1466 QKEARSLSTEL-----FLKNAYEESLEHETFPKENKNNLEISDLT---EQLGASQSKI 1518
 QY 449 QRLKCLALQTERNDLNKRVQDLQA 474
 Db 1519 HELEKVRKQLDAEKLELQAALEBAEA 1544

RESULT 27

S28261
 centromere protein E - human
 N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E
 C:Species: Homo sapiens (man)
 C>Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
 C:Accession: S28261
 R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
 Nature 359, 536-539, 1992
 A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
 A:Reference number: S28261; MUID:93024922; PMID:1406971
 A:Accession: S28261
 A:Molecule type: mRNA
 A:Residues: 1-2663 <YEN>
 A:Cross-references: EMBL:Z15005; NID:929864; PIDN:CAA78727.1; PID:929865
 C:Genetics:
 A:Gene: GDB:CENPE
 A:Cross-references: GDB:361164; OMIM:117143
 A:Map position: 4q24-q25
 C:Superfamily: centromere protein E; kinesin motor domain homology
 C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
 P:7-335/Domain: kinesin motor domain homology <MMOT>
 P:86-93/Region: nucleotide-binding motif A (P-loop)

1

QY 171 EKALCKKYBALHEHRSQOM--KLQKQSQVLVQKDHRLHGHSHKAVLARSLSL 228
 DB 2428 QEAARL-----RLAERDLAQORALAEKMLKERM-QAVQEAATLKAEE----- 2468
 QY 229 CRELQHNESLKEGVORAREEERK-----EVTSHFQVTL--NDIQLOMBOHNRNSK 281
 DB 2469 -ALIQOQKELAQEQARRLOADKEQVQAQVIVETQGFQTLAERQRLKESAEAR-LK 2526
 QY 282 LRQENN-----ELAEKLLKLEQVYHLEREHDKVFKHDLQOQVLDAKLQQAQ---E 330
 DB 2527 LRMAESRAQAAEEDAQFRQAEBIGELKHRLTELATQKVTQVTLQTLBIQORQSDQDAE 2586
 QY 331 MLKBAERHQRKDFLLKAVESORCELMK--QOETHLKOQALAEYKFEFQNTLSKS 388
 DB 2587 RUREATAERERKLEKQBAKULQLSXSEKQVQEQIIOETQALQKSFSEKSDLSLQRE 2646
 QY 389 SEVFTTFKQEMERNTK-----KIKLEKETMYRSRWESSNKALLEMAEEXTVRDKEL 441
 DB 2647 RFI-----EQEKAKLEQIFQDEVAKAKQLQBEQQRQOQOQMEQEQELVASMEEARRRQREA 2702
 QY 442 -EGL-----QVKIQLEKLC-----RALQTERNDLNKRVODLS-----ACQCSL 480
 DB 2703 EGVREKQELQLEQOQOQKELAEENQRLERLQRLERHEHRAALAHSEIATSQAAA 2762
 QY 481 TDSGPRERP--EGPGQA 496
 DB 2763 TKALPNGRDALDGPMEAR 2780
 RESULT 32
 T13030
 microtubule binding protein D-CLIP-190 - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: T13030
 R:Lantz, V.A.; Miller, K.G.
 J. Cell Biol. 140, 897-910, 1998
 A:Title: A class VI unconventional myosin is associated with a homologue of a microtubul
 A:Reference number: Z17588; MUID:98139549; PMID:9472041
 A:Accession: T13030
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1690 <LAN>
 A:Cross-references: EMBL:AF041382; NID:G2773362; PID:G2773363; PIDN:AAB96783.1
 A:Experimental source: strain Oregon R
 C:Genetics:
 A:Cross-references: FlyBase:FBgn0020503
 C:Keywords: cytoskeleton
 Query Match 8.8%; Score 237.5; DB 2; Length 1690;
 Best Local Similarity 23.1%; Pred. No. 0.0079;
 Matches 126; Conservative 104; Mismatches 178; Indels 137; Gaps 24;
 QY 33 SQAPKPEGQAQTAQSGALRDVSELSROLEDILSTYCVVDNQGPGEDGAQGEPAEPE 92
 DB 352 SSTPVKPIILATPKSQFS--KQDLREKQKHQVKEKLVREDDLK-----E 392
 QY 93 DAERSRTYVARNGEPFPTPVVYGEKPSKGDPNTEIRQS-DEV---GDRHRRPQ-EKK 147
 DB 393 DAQNOALQLOQNIQNELKARIVELESALGNERKTEHLCSDIDEAQFCQGBELNAQSQVYKE 452
 QY 148 KAKGLGKEITLLMQ---TINTLSTP-----EKLALCKK-----YAEEL 184
 DB 453 KIHLESKITKLVSATPSLSQSLPDLPSDDGALQEQEIAQLOEQNTTCQKEVBSRIABQL 512
 QY 185 BEHNSQOMKLLQCK-----QSQVLQVQKDHRLHGHSHKAVLARSLSLCHRELQHRNLSX 240
 DB 513 EEOQLRENVKYLNEQIATLQSELVSKDEAL-----EKFSLECCGIENLREL-----ALLK 564
 QY 241 EGVQVAREE-----EKKKEV-----TSHFQVTLNDI 268
 DB 565 BENKQAEQAEAEFRKLAESKVEFLRLSSLQNLKATSDSLESERVNKSDECEILOTEV 624

QY 269 QLOMEOHNRNSKL-----ROENMELAE-----RLKCLIEQVYELREHIDKVPFKHKL 316
 DB 625 NMRDEQIRELQQLDEVTTLQNVQKADSSALDDMLRLOK--EGTEEKSTLEKTI----- 676
 QY 317 QOQLVADKLQQAQEMLEKEAERHQRKDFLLKAVESQRMCM-----ELMKQOETHL 367
 DB 677 EKELVQIK-EQAATLQDKQLEKQISD--LKQLAEQKLVKREKTENAINQIQLEKES-I 732
 QY 368 KOALALYTEKEFEFQNTLSKSSEVFTTFKQ-----EMEKMTKKIKLEKETMYRSRW 420
 DB 733 EQALAKQNELEDFKQKQSSSEVHLQELKAQNTOKDLVELVESGSLKLLQOOLEKTLGH 792
 QY 421 ESSNKALLEMAEEXTV-----RDKELEGVQVKIQLEKLCRALQTERNDLNKRVQDLSAGG 476
 DB 793 EKLQAALBELKKEKSTIIKEKEQELQQLQSKSAESSEALKVVQVLEQLQO--QAAASGE 850
 QY 477 QQSULT 481
 DB 851 EGSKT 855
 RESULT 33
 I48153
 myosin heavy chain beta, cardiac muscle [similarity] - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Feb-2001
 C:Accession: I48153; A28298
 R:Wang, R.; Sole, M.J.; Cukerman, E.; Liew, C.C.
 J. Mol. Cell. Cardiol. 26, 1155-1165, 1994
 A:Title: Characterization and nucleotide sequence of the cardiac alpha-myosin heavy cha
 A:Reference number: I48153; MUID:95115033; PMID:7815459
 A:Accession: I48153
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1934 <RES>
 A:Cross-references: GB:L12104; NID:9402371; PIDN:AAA62313.1; PID:9402372
 R:Jandreski, M.A.; Sole, M.J.; Liew, C.C.
 Nucleic Acids Res. 16, 4737, 1988
 A:Title: Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin heavy chain.
 A:Reference number: A28298; MUID:88247788; PMID:3380703
 A:Accession: A28298
 A:Molecule type: mRNA
 A:Residues: 962-965, 'E', 967-980, 'E', 981-985, 'Q', 987-1007, 'A', 1009, 'E', 1011, 'RKT', 1015-1
 536, 'L', 1538-1555, 'K', 1557-1934 <JAN>
 A:Cross-references: GB:IX07273; NID:949640; PIDN:CAA30256.1; PID:949641
 A>Note: the authors translated the codon GTG for residue 1504 as leu
 C:Genetics:
 A:Introns: 66/3; 114/3; 167/1; 176/2; 212/3; 243/3; 265/1; 298/1; 332/3; 379/1; 418/3; 4
 23/3; 1389/2; 1450/3; 1506/1; 1547/3; 1650/3; 1718/3; 1760/3; 1852/3; 1884/3; 1929/3
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; muscle; nucleotide I
 F; 87-765/Domain: myosin motor domain homology <WNOT>
 F; 177-184/Region: nucleotide-binding motif A (P-loop)
 Query Match 8.8%; Score 237; DB 2; Length 1934;
 Best Local Similarity 23.3%; Pred. No. 0.0096;
 Matches 129; Conservative 88; Mismatches 208; Indels 128; Gaps 20;
 QY 12 EGAQRPQAAVAEAGGSGQAQKPEGAQTAQSGALRDVSELSROLEDILSTYC 71
 DB 1034 EGSLEQEKVRMDLE-----RAKKLEGDLKLTQES--IMDLND-KQDLDEKLKKD 1083
 QY 72 VDNQGGPGEDGAQGEPAEPEDA-----EKSRTYVAR-----NGEPFPTPVVYGEKBP 119
 DB 1084 FELN-----ALNARIDEQALGSLQKLKELQARIEELSELEAERTAKVEKL 1135
 QY 120 SKGDPNTEIRQS-DEVG-----DRHRRPQKKAQKGLKEITLLQTLTLPTEE 171
 DB 1136 SDSLSELEISELEAGGATSVQIEMNKKGREAFQMRDLSEATL-----QHEA 1186
 QY 172 KLALCKKY-----AELLBEHNSQOMKLLQKQSQVLQVQKDHRLHGHSHKAVLARSLSL 227

Db 1187 TAAALRKHADSVABLGEQIDNLQVRQKLEKSEFQLELDVNTSMEOI IKANLEK 1246
Qy 228 LCR-----ELQRN-----RLKBE-----241
Db 1247 MCRLEPDMNHRKSAEETQSYNDLTSQRAKLTQNGELSRQDKEALISQLTGKLT 1306
Qy 242 -----EGVQAREREKKEVTSF-FOVTLNDLQLOMEQNNRNSK-----LRQENME 288
Db 1307 YTOQLEDLKRQLEBEVKAQNTLAHALQASRHDCLLEQYEEETEAARAEQCVLSKANSE 1366
Qy 289 LAERLKKLIEQYELREBEHIDVFKHDKLQOQLVDKQLQAOE-----MLKBAEERHOR 341
Db 1367 VAQWRKYETDAIOITELSEA--KKLAGLQDA--EBAEAVNAKCSSLEKTKHRLQN 1422
Qy 342 EKDFLKEAVESORMCELMKQOETHLKKQALYTEKPEEFQNTLSKSEVFTTFKQMEK 401
Db 1423 EIEDLVVDVERNSAAAALDKRQNFQKILAEKQKYEESQSELESSQKEARSLSLTEL-- 1480
Qy 402 MTKIKKLEKETTMYRSWESSNKALLEMAEKTVRDKLEGLQVKIORLEKLCRALQTE 461
Db 1481 --FKKNAYESLHLETFKRNKNLQBEISDLT---EQLGSTGKSITHELEKTRKQLEAE 1535
Qy 462 RNDLKNRVQDLSA 474
Db 1536 KMELOQSALEAEAA 1548
RESULT 34
152300
gi|152300|g|giantin - human
N:Altemate names: gcp372
C:Species: Homo sapiens (man)
C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C:Accession: I52300
R:Shonda, M.; Misumi, Y.; Fujiwara, T.; Nishioka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in
A:Reference number: I52300; MUID:95100974; PMID:7802676
C:Accession: I52300
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3225 <RS>
A:Cross-references: GB:D25542; NID:9662389; PIDN:BA05025.1; PID:9808869
C:Superfamily: Giantin

Query Match 8.8%; Score 236.5; DB 2; Length 3225;
Best Local Similarity 22.8%; Pred. No. 0.017;
Matches 119; Conservative 99; Mismatches 178; Indels 125; Gaps 18;
Qy 52 LADVSELSRQLEDILSTYCYDNNQGGPGEDGAQCEPAEDPAEKSRITYVARNGEPEPTP 111
Db 2248 LQGENKELLSQLETRHLY-----HSSQNELAKLESELKSL-----KQQLTD 2289
Qy 112 VYVGEKPSKGPNTPEIRQSDVGDVDRHRRPQE-----KKKAGLGKGIITLLMQT 162
Db 2290 LNSLEKCKEQGNLEGIIRQEQADIQNSKESYEQLETDQASRLTSLAEELINMKQK 2349
Qy 163 LNTLSTPEKLAALCKKVAELLEHRNSQKMLQKQSQQLVQKQHLRGHSHKAVLAR 222
Db 2350 IISLLSGKEE--AIQVIAELRQOH--DKETKELENLLSOEEENIVLEENKAV---2401
Qy 223 SKLESICRELQRHNSLKEEGYQARAREBEKKEVTS-----HFQVT 264
Db 2402 DKTNGLMETL-----KTIKENIQQKQALDSFVMSSSLNDNRDRIVGYQQLREHLSII 2457
Qy 265 LNDIOLQMEQHNRNSKLQOE-----NMELAEKLLI1EQYELRESHIDKVFKH 313
Db 2458 LEKDLQIQEAAE--NNKLEERIGLRSHMDLNSNAKLDABLI-QY---REDLQNVITI 2512
Qy 314 KILQO--QLVDKQLQAOEM-----LKABERH-----QREKDFLKEAVE 352
Db 2513 KDSQQLLEVLQQLQKLENLKYNKYLSEKLSESEANEDLRRSFNALQOEKQDLSKETES 2572

Qy 353 SQRMCELMKQOETHLQO--LALYTEKPEEFQNTLSKSEVFTTFKQEM-----399
Db 2573 LKVSISQLTQVTAQOEGTGLYHAQKLVKEEVRHLSALFSSQKRIAELEELVQVQ 2632
Qy 400 -----EKMTKKIKKLEKETTMYRSWESSN-----KALLEMAEKTVRDKEL 441
Db 2633 KEAAKKGVBIEKDKLKKELKHLHDAGIMNETAABERVABELARDLVEMEOKLLAVTKEN 2692
Qy 442 EQLQVKIQLKLEKLCRALQTERNDLKNRVODLSAGGQSLTD 482
Db 2693 KGLTAQIQSGRSMGSLQSRDHANEELDELCKKYDASLKE 2733
RESULT 35
A70387
Conserved hypothetical protein aq_1006 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Feb-2001
C:Accession: A70387
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
C:Accession: A70387
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-978 <AQF>
A:Cross-references: GB:AE000718; NID:g2983504; PIDN:AAC07092.1; PID:g2983515; GB:AB0006
A:Experimental source: strain V75
C:Genetics:
C:Gene: aq_1006
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 8.7%; Score 236; DB 2; Length 978;
Best Local Similarity 21.1%; Pred. No. 0.0052;
Matches 107; Conservative 103; Mismatches 180; Indels 118; Gaps 17;
Qy 14 AQERPOAPVABEAGPGSSQAPRKPEGAQAPTAQSGALRDVSEELSRQLEDILSTYCYD 73
Db 314 AQBELNR-----IEAEKEKFEKEREKLEHRLKQLQKIQEIKELKELUS-QUSSLK-----363
Qy 74 NQGGPGEDGAQCEPAEDPAEKSRITYVARNGEPEPTP--VYVGEKPSKGPNTPEIRQ 131
Db 364 -----EKERYEQAKQEPEDLSERVEKGGKLVASTEEKLEIKE 402
Qy 132 --SDEVDGRDRRPOEKKKAGKIGKITLLMQTLNTLSTPEKLAALCKKVAELLEHRN 189
Db 403 LFSEE-----EYTSLKMKERLIVELQRLKELKEGQLENLTQYKKEKKVHEK 452
Qy 190 SQKMKLLQKQSQQLVQEKDHLRGHSHKAVLARSKLESICRELQRH--RSLKEEGVQRA 247
Db 453 VLNELKELR---ELKERELHYAHVAVSYLSFGDTCPVCGGIYRGKALENVDAEGISEL 509
Qy 248 REEEKKEKRYTSHFQVTL-----NDIQLOMSQHNERNKSLRQENME-LAEELKLLIEQ 299
Db 510 KHAKELKEBEREDITTLKLVYAQKINSLEKEEMKLENEVEELKEITPENLKERIKGL-BE 568
Qy 300 YELREEHID-----KVFKHDLQOQLVDKQLQAOE-----EMLKEAERHOREKDFLLKE 349
Db 569 LRTEKLEHKLKYNKALSDROKKEBAQKLAQTELELKE-----KIREKSLVKE 624
Qy 350 AVESQRCMLMKQOETHLQOALYTEKPEEFQNTLSKSEVFTTFKQEMQNTKKI KKL 409
Db 625 FKELRYV-----ERLEDYBESLKEEINYNSKLQIEBEKEKLLRKH 665
Qy 410 EKETTMYRSWESS-----NKALLEMAEKTVRDKELG-----443
Db 666 FEELSSRSKSLGELSALNESINSLEERKEKLELANIYEVAKSPREVELYLGDEAE 725
Qy 444 LQVKIORLEKLCRALQTERNDLKNRVQD 471
Db 726 LERKLEFEESFQSLKAKKSEIEKLE 753

A;Residues: 1-1939 <MAT>
A;Cross-references: DDBJ:D00943; NID:g219523; PIDN:BA00791.1; PID:g219524
A;Accession: B46752
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1461 <MA2>
R;Epp, T.A.; Dixon, I.M.C.; Wang, H.Y.; Sole, M.J.; Liew, C.C.
Genomics 16, 505-509, 1993
A;Title: Structural organization of the human cardiac alpha-myosin heavy chain gene (MYH7)
A;Reference number: A49354; MUID:94140346; PMID:8307559
A;Accession: A49354
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-87, 'Q', 89-573, 'Q', 575-607, 'A', 609-743, 'T', 745-789, 'M', 791-1013, 'V', 1015-1019, 'E', 1021-1023, 'D', 1025-1027, 'S', 1029-1031, 'P', 1033-1035, 'L', 1037-1039, 'K', 1041-1043, 'R', 1045-1047, 'G', 1049-1051, 'A', 1053-1055, 'C', 1057-1059, 'T', 1061-1063, 'G', 1065-1067, 'C', 1069-1071, 'A', 1073-1075, 'G', 1077-1079, 'C', 1081-1083, 'T', 1085-1087, 'A', 1089-1091, 'C', 1093-1095, 'G', 1097-1099, 'C', 1101-1103, 'T', 1105-1107, 'A', 1109-1111, 'C', 1113-1115, 'G', 1117-1119, 'C', 1121-1123, 'T', 1125-1127, 'A', 1129-1131, 'C', 1133-1135, 'G', 1137-1139, 'C', 1141-1143, 'T', 1145-1147, 'A', 1149-1151, 'C', 1153-1155, 'G', 1157-1159, 'C', 1161-1163, 'T', 1165-1167, 'A', 1169-1171, 'C', 1173-1175, 'G', 1177-1179, 'C', 1181-1183, 'T', 1185-1187, 'A', 1189-1191, 'C', 1193-1195, 'G', 1197-1199, 'C', 1201-1203, 'T', 1205-1207, 'A', 1209-1211, 'C', 1213-1215, 'G', 1217-1219, 'C', 1221-1223, 'T', 1225-1227, 'A', 1229-1231, 'C', 1233-1235, 'G', 1237-1239, 'C', 1241-1243, 'T', 1245-1247, 'A', 1249-1251, 'C', 1253-1255, 'G', 1257-1259, 'C', 1261-1263, 'T', 1265-1267, 'A', 1269-1271, 'C', 1273-1275, 'G', 1277-1279, 'C', 1281-1283, 'T', 1285-1287, 'A', 1289-1291, 'C', 1293-1295, 'G', 1297-1299, 'C', 1301-1303, 'T', 1305-1307, 'A', 1309-1311, 'C', 1313-1315, 'G', 1317-1319, 'C', 1321-1323, 'T', 1325-1327, 'A', 1329-1331, 'C', 1333-1335, 'G', 1337-1339, 'C', 1341-1343, 'T', 1345-1347, 'A', 1349-1351, 'C', 1353-1355, 'G', 1357-1359, 'C', 1361-1363, 'T', 1365-1367, 'A', 1369-1371, 'C', 1373-1375, 'G', 1377-1379, 'C', 1381-1383, 'T', 1385-1387, 'A', 1389-1391, 'C', 1393-1395, 'G', 1397-1399, 'C', 1401-1403, 'T', 1405-1407, 'A', 1409-1411, 'C', 1413-1415, 'G', 1417-1419, 'C', 1421-1423, 'T', 1425-1427, 'A', 1429-1431, 'C', 1433-1435, 'G', 1437-1439, 'C', 1441-1443, 'T', 1445-1447, 'A', 1449-1451, 'C', 1453-1455, 'G', 1457-1459, 'C', 1461-1463, 'T', 1465-1467, 'A', 1469-1471, 'C', 1473-1475, 'G', 1477-1479, 'C', 1481-1483, 'T', 1485-1487, 'A', 1489-1491, 'C', 1493-1495, 'G', 1497-1499, 'C', 1501-1503, 'T', 1505-1507, 'A', 1509-1511, 'C', 1513-1515, 'G', 1517-1519, 'C', 1521-1523, 'T', 1525-1527, 'A', 1529-1531, 'C', 1533-1535, 'G', 1537-1539, 'C', 1541-1543, 'T', 1545-1547, 'A', 1549-1551, 'C', 1553-1555, 'G', 1557-1559, 'C', 1561-1563, 'T', 1565-1567, 'A', 1569-1571, 'C', 1573-1575, 'G', 1577-1579, 'C', 1581-1583, 'T', 1585-1587, 'A', 1589-1591, 'C', 1593-1595, 'G', 1597-1599, 'C', 1601-1603, 'T', 1605-1607, 'A', 1609-1611, 'C', 1613-1615, 'G', 1617-1619, 'C', 1621-1623, 'T', 1625-1627, 'A', 1629-1631, 'C', 1633-1635, 'G', 1637-1639, 'C', 1641-1643, 'T', 1645-1647, 'A', 1649-1651, 'C', 1653-1655, 'G', 1657-1659, 'C', 1661-1663, 'T', 1665-1667, 'A', 1669-1671, 'C', 1673-1675, 'G', 1677-1679, 'C', 1681-1683, 'T', 1685-1687, 'A', 1689-1691, 'C', 1693-1695, 'G', 1697-1699, 'C', 1701-1703, 'T', 1705-1707, 'A', 1709-1711, 'C', 1713-1715, 'G', 1717-1719, 'C', 1721-1723, 'T', 1725-1727, 'A', 1729-1731, 'C', 1733-1735, 'G', 1737-1739, 'C', 1741-1743, 'T', 1745-1747, 'A', 1749-1751, 'C', 1753-1755, 'G', 1757-1759, 'C', 1761-1763, 'T', 1765-1767, 'A', 1769-1771, 'C', 1773-1775, 'G', 1777-1779, 'C', 1781-1783, 'T', 1785-1787, 'A', 1789-1791, 'C', 1793-1795, 'G', 1797-1799, 'C', 1801-1803, 'T', 1805-1807, 'A', 1809-1811, 'C', 1813-1815, 'G', 1817-1819, 'C', 1821-1823, 'T', 1825-1827, 'A', 1829-1831, 'C', 1833-1835, 'G', 1837-1839, 'C', 1841-1843, 'T', 1845-1847, 'A', 1849-1851, 'C', 1853-1855, 'G', 1857-1859, 'C', 1861-1863, 'T', 1865-1867, 'A', 1869-1871, 'C', 1873-1875, 'G', 1877-1879, 'C', 1881-1883, 'T', 1885-1887, 'A', 1889-1891, 'C', 1893-1895, 'G', 1897-1899, 'C', 1901-1903, 'T', 1905-1907, 'A', 1909-1911, 'C', 1913-1915, 'G', 1917-1919, 'C', 1921-1923, 'T', 1925-1927, 'A', 1929-1931, 'C', 1933-1935, 'G', 1937-1939, 'C', 1941-1943, 'T', 1945-1947, 'A', 1949-1951, 'C', 1953-1955, 'G', 1957-1959, 'C', 1961-1963, 'T', 1965-1967, 'A', 1969-1971, 'C', 1973-1975, 'G', 1977-1979, 'C', 1981-1983, 'T', 1985-1987, 'A', 1989-1991, 'C', 1993-1995, 'G', 1997-1999, 'C', 2001-2003, 'T', 2005-2007, 'A', 2009-2011, 'C', 2013-2015, 'G', 2017-2019, 'C', 2021-2023, 'T', 2025-2027, 'A', 2029-2031, 'C', 2033-2035, 'G', 2037-2039, 'C', 2041-2043, 'T', 2045-2047, 'A', 2049-2051, 'C', 2053-2055, 'G', 2057-2059, 'C', 2061-2063, 'T', 2065-2067, 'A', 2069-2071, 'C', 2073-2075, 'G', 2077-2079, 'C', 2081-2083, 'T', 2085-2087, 'A', 2089-2091, 'C', 2093-2095, 'G', 2097-2099, 'C', 2101-2103, 'T', 2105-2107, 'A', 2109-2111, 'C', 2113-2115, 'G', 2117-2119, 'C', 2121-2123, 'T', 2125-2127, 'A', 2129-2131, 'C', 2133-2135, 'G', 2137-2139, 'C', 2141-2143, 'T', 2145-2147, 'A', 2149-2151, 'C', 2153-2155, 'G', 2157-2159, 'C', 2161-2163, 'T', 2165-2167, 'A', 2169-2171, 'C', 2173-2175, 'G', 2177-2179, 'C', 2181-2183, 'T', 2185-2187, 'A', 2189-2191, 'C', 2193-2195, 'G', 2197-2199, 'C', 2201-2203, 'T', 2205-2207, 'A', 2209-2211, 'C', 2213-2215, 'G', 2217-2219, 'C', 2221-2223, 'T', 2225-2227, 'A', 2229-2231, 'C', 2233-2235, 'G', 2237-2239, 'C', 2241-2243, 'T', 2245-2247, 'A', 2249-2251, 'C', 2253-2255, 'G', 2257-2259, 'C', 2261-2263, 'T', 2265-2267, 'A', 2269-2271, 'C', 2273-2275, 'G', 2277-2279, 'C', 2281-2283, 'T', 2285-2287, 'A', 2289-2291, 'C', 2293-2295, 'G', 2297-2299, 'C', 2301-2303, 'T', 2305-2307, 'A', 2309-2311, 'C', 2313-2315, 'G', 2317-2319, 'C', 2321-2323, 'T', 2325-2327, 'A', 2329-2331, 'C', 2333-2335, 'G', 2337-2339, 'C', 2341-2343, 'T', 2345-2347, 'A', 2349-2351, 'C', 2353-2355, 'G', 2357-2359, 'C', 2361-2363, 'T', 2365-2367, 'A', 2369-2371, 'C', 2373-2375, 'G', 2377-2379, 'C', 2381-2383, 'T', 2385-2387, 'A', 2389-2391, 'C', 2393-2395, 'G', 2397-2399, 'C', 2401-2403, 'T', 2405-2407, 'A', 2409-2411, 'C', 2413-2415, 'G', 2417-2419, 'C', 2421-2423, 'T', 2425-2427, 'A', 2429-2431, 'C', 2433-2435, 'G', 2437-2439, 'C', 2441-2443, 'T', 2445-2447, 'A', 2449-2451, 'C', 2453-2455, 'G', 2457-2459, 'C', 2461-2463, 'T', 2465-2467, 'A', 2469-2471, 'C', 2473-2475, 'G', 2477-2479, 'C', 2481-2483, 'T', 2485-2487, 'A', 2489-2491, 'C', 2493-2495, 'G', 2497-2499, 'C', 2501-2503, 'T', 2505-2507, 'A', 2509-2511, 'C', 2513-2515, 'G', 2517-2519, 'C', 2521-2523, 'T', 2525-2527, 'A', 2529-2531, 'C', 2533-2535, 'G', 2537-2539, 'C', 2541-2543, 'T', 2545-2547, 'A', 2549-2551, 'C', 2553-2555, 'G', 2557-2559, 'C', 2561-2563, 'T', 2565-2567, 'A', 2569-2571, 'C', 2573-2575, 'G', 2577-2579, 'C', 2581-2583, 'T', 2585-2587, 'A', 2589-2591, 'C', 2593-2595, 'G', 2597-2599, 'C', 2601-2603, 'T', 2605-2607, 'A', 2609-2611, 'C', 2613-2615, 'G', 2617-2619, 'C', 2621-2623, 'T', 2625-2627, 'A', 2629-2631, 'C', 2633-2635, 'G', 2637-2639, 'C', 2641-2643, 'T', 2645-2647, 'A', 2649-2651, 'C', 2653-2655, 'G', 2657-2659, 'C', 2661-2663, 'T', 2665-2667, 'A', 2669-2671, 'C', 2673-2675, 'G', 2677-2679, 'C', 2681-2683, 'T', 2685-2687, 'A', 2689-2691, 'C', 2693-2695, 'G', 2697-2699, 'C', 2701-2703, 'T', 2705-2707, 'A', 2709-2711, 'C', 2713-2715, 'G', 2717-2719, 'C', 2721-2723, 'T', 2725-2727, 'A', 2729-2731, 'C', 2733-2735, 'G', 2737-2739, 'C', 2741-2743, 'T', 2745-2747, 'A', 2749-2751, 'C', 2753-2755, 'G', 2757-2759, 'C', 2761-2763, 'T', 2765-2767, 'A', 2769-2771, 'C', 2773-2775, 'G', 2777-2779, 'C', 2781-2783, 'T', 2785-2787, 'A', 2789-2791, 'C', 2793-2795, 'G', 2797-2799, 'C', 2801-2803, 'T', 2805-2807, 'A', 2809-2811, 'C', 2813-2815, 'G', 2817-2819, 'C', 2821-2823, 'T', 2825-2827, 'A', 2829-2831, 'C', 2833-2835, 'G', 2837-2839, 'C', 2841-2843, 'T', 2845-2847, 'A', 2849-2851, 'C', 2853-2855, 'G', 2857-2859, 'C', 2861-2863, 'T', 2865-2867, 'A', 2869-2871, 'C', 2873-2875, 'G', 2877-2879, 'C', 2881-2883, 'T', 2885-2887, 'A', 2889-2891, 'C', 2893-2895, 'G', 2897-2899, 'C', 2901-2903, 'T', 2905-2907, 'A', 2909-2911, 'C', 2913-2915, 'G', 2917-2919, 'C', 2921-2923, 'T', 2925-2927, 'A', 2929-2931, 'C', 2933-2935, 'G', 2937-2939, 'C', 2941-2943, 'T', 2945-2947, 'A', 2949-2951, 'C', 2953-2955, 'G', 2957-2959, 'C', 2961-2963, 'T', 2965-2967, 'A', 2969-2971, 'C', 2973-2975, 'G', 2977-2979, 'C', 2981-2983, 'T', 2985-2987, 'A', 2989-2991, 'C', 2993-2995, 'G', 2997-2999, 'C', 3001-3003, 'T', 3005-3007, 'A', 3009-3011, 'C', 3013-3015, 'G', 3017-3019, 'C', 3021-3023, 'T', 3025-3027, 'A', 3029-3031, 'C', 3033-3035, 'G', 3037-3039, 'C', 3041-3043, 'T', 3045-3047, 'A', 3049-3051, 'C', 3053-3055, 'G', 3057-3059, 'C', 3061-3063, 'T', 3065-3067, 'A', 3069-3071, 'C', 3073-3075, 'G', 3077-3079, 'C', 3081-3083, 'T', 3085-3087, 'A', 3089-3091, 'C', 3093-3095, 'G', 3097-3099, 'C', 3101-3103, 'T', 3105-3107, 'A', 3109-3111, 'C', 3113-3115, 'G', 3117-3119, 'C', 3121-3123, 'T', 3125-3127, 'A', 3129-3131, 'C', 3133-3135, 'G', 3137-3139, 'C', 3141-3143, 'T', 3145-3147, 'A', 3149-3151, 'C', 3153-3155, 'G', 3157-3159, 'C', 3161-3163, 'T', 3165-3167, 'A', 3169-3171, 'C', 3173-3175, 'G', 3177-3179, 'C', 3181-3183, 'T', 3185-3187, 'A', 3189-3191, 'C', 3193-3195, 'G', 3197-3199, 'C', 3201-3203, 'T', 3205-3207, 'A', 3209-3211, 'C', 3213-3215, 'G', 3217-3219, 'C', 3221-3223, 'T', 3225-3227, 'A', 3229-3231, 'C', 3233-3235, 'G', 3237-3239, 'C', 3241-3243, 'T', 3245-3247, 'A', 3249-3251, 'C', 3253-3255, 'G', 3257-3259, 'C', 3261-3263, 'T', 3265-3267, 'A', 3269-3271, 'C', 3273-3275, 'G', 3277-3279, 'C', 3281-3283, 'T', 3285-3287, 'A', 3289-3291, 'C', 3293-3295, 'G', 3297-3299, 'C', 3301-3303, 'T', 3305-3307, 'A', 3309-3311, 'C', 3313-3315, 'G', 3317-3319, 'C', 3321-3323, 'T', 3325-3327, 'A', 3329-3331, 'C', 3333-3335, 'G', 3337-3339, 'C', 3341-3343, 'T', 3345-3347, 'A', 3349-3351, 'C', 3353-3355, 'G', 3357-3359, 'C', 3361-3363, 'T', 3365-3367, 'A', 3369-3371, 'C', 3373-3375, 'G', 3377-3379, 'C', 3381-3383, 'T', 3385-3387, 'A', 3389-3391, 'C', 3393-3395, 'G', 3397-3399, 'C', 3401-3403, 'T', 3405-3407, 'A', 3409-3411, 'C', 3413-3415, 'G', 3417-3419, 'C', 3421-3423, 'T', 3425-3427, 'A', 3429-3431, 'C', 3433-3435, 'G', 3437-3439, 'C', 3441-3443, 'T', 3445-3447, 'A', 3449-3451, 'C', 3453-3455, 'G', 3457-3459, 'C', 3461-3463, 'T', 3465-3467, 'A', 3469-3471, 'C', 3473-3475, 'G', 3477-3479, 'C', 3481-3483, 'T', 3485-3487, 'A', 3489-3491, 'C', 3493-3495, 'G', 3497-3499, 'C', 3501-3503, 'T', 3505-3507, 'A', 3509-3511, 'C', 3513-3515, 'G', 3517-3519, 'C', 3521-3523, 'T', 3525-3527, 'A', 3529-3531, 'C', 3533-3535, 'G', 3537-3539, 'C', 3541-3543, 'T', 3545-3547, 'A', 3549-3551, 'C', 3553-3555, 'G', 3557-3559, 'C', 3561-3563, 'T', 3565-3567, 'A', 3569-3571, 'C', 3573-3575, 'G', 3577-3579, 'C', 3581-3583, 'T', 3585-3587, 'A', 3589-3591, 'C', 3593-3595, 'G', 3597-3599, 'C', 3601-3603, 'T', 3605-3607, 'A', 3609-3611, 'C', 3613-3615, 'G', 3617-3619, 'C', 3621-3623, 'T', 3625-3627, 'A', 3629-3631, 'C', 3633-3635, 'G', 3637-3639, 'C', 3641-3643, 'T', 3645-3647, 'A', 3649-3651, 'C', 3653-3655, 'G', 3657-3659, 'C', 3661-3663, 'T', 3665-3667, 'A', 3669-3671, 'C', 3673-3675, 'G', 3677-3679, 'C', 3681-3683, 'T', 3685-3687, 'A', 3689-3691, 'C', 3693-3695, 'G', 3697-3699, 'C', 3701-3703, 'T', 3705-3707, 'A', 3709-3711, 'C', 3713-3715, 'G', 3717-3719, 'C', 3721-3723, 'T', 3725-3727, 'A', 3729-3731, 'C', 3733-3735, 'G', 3737-3739, 'C', 3741-3743, 'T', 3745-3747, 'A', 3749-3751, 'C', 3753-3755, 'G', 3757-3759, 'C', 3761-3763, 'T', 3765-3767, 'A', 3769-3771, 'C', 3773-3775, 'G', 3777-3779, 'C', 3781-3783, 'T', 3785-3787, 'A', 3789-3791, 'C', 3793-3795, 'G', 3797-3799, 'C', 3801-3803, 'T', 3805-3807, 'A', 3809-3811, 'C', 3813-3815, 'G', 3817-3819, 'C', 3821-3823, 'T', 3825-3827, 'A', 3829-3831, 'C', 3833-3835, 'G', 3837-3839, 'C', 3841-3843, 'T', 3845-3847, 'A', 3849-3851, 'C', 3853-3855, 'G', 3857-3859, 'C', 3861-3863, 'T', 3865-3867, 'A', 3869-3871, 'C', 3873-3875, 'G', 3877-3879, 'C', 3881-3883, 'T', 3885-3887, 'A', 3889-3891, 'C', 3893-3895, 'G', 3897-3899, 'C', 3901-3903, 'T', 3905-3907, 'A', 3909-3911, 'C', 3913-3915, 'G', 3917-3919, 'C', 3921-3923, 'T', 3925-3927, 'A', 3929-3931, 'C', 3933-3935, 'G', 3937-3939, 'C', 3941-3943, 'T', 3945-3947, 'A', 3949-3951, 'C', 3953-3955, 'G', 3957-3959, 'C', 3961-3963, 'T', 3965-3967, 'A', 3969-3971, 'C', 3973-3975, 'G', 3977-3979, 'C', 3981-3983, 'T', 3985-3987, 'A', 3989-3991, 'C', 3993-3995, 'G', 3997-3999, 'C', 4001-4003, 'T', 4005-4007, 'A', 4009-4011, 'C', 4013-4015, 'G', 4017-4019, 'C', 4021-4023, 'T', 4025-4027, 'A', 4029-4031, 'C', 4033-4035, 'G', 4037-4039, 'C', 4041-4043, 'T', 4045-4047, 'A', 4049-4051, 'C', 4053-4055, 'G', 4057-4059, 'C', 4061-4063, 'T', 4065-4067, 'A', 4069-4071, 'C', 4073-4075, 'G', 4077-4079, 'C', 4081-4083, 'T', 4085-4087, 'A', 4089-4091, 'C', 4093-4095, 'G', 4097-4099, 'C', 4101-4103, 'T', 4105-4107, 'A', 4109-4111, 'C', 4113-4115, 'G', 4117-4119, 'C', 4121-4123, 'T', 4125-4127, 'A', 4129-4131, 'C', 4133-4135, 'G', 4137-4139, 'C', 4141-4143, 'T', 4145-4147, 'A', 4149-4151, 'C', 4153-4155, 'G', 4157-4159, 'C', 4161-4163, 'T', 4165-4167, 'A', 4169-4171, 'C', 4173-4175, 'G', 4177-4179, 'C', 4181-4183, 'T', 4185-4187, 'A', 4189-4191, 'C', 4193-4195, 'G', 4197-4199, 'C', 4201-4203, 'T', 4205-4207, 'A', 4209-4211, 'C', 4213-4215, 'G', 4217-4219, 'C', 4221-4223, 'T', 4225-4227, 'A', 4229-4231, 'C', 4233-4235, 'G', 4237-4239, 'C', 4241-4243, 'T', 4245-4247, 'A', 4249-4251, 'C', 4253-4255, 'G', 4257-4259, 'C', 4261-4263, 'T', 4265-4267, 'A', 4269-4271, 'C', 4273-4275, 'G', 4277-4279, 'C', 4281-4283, 'T', 4285-4287, 'A', 4289-4291, 'C', 4293-4295, 'G', 4297-4299, 'C', 4301-4303, 'T', 4305-4307, 'A', 4309-4311, 'C', 4313-4315, 'G', 4317-4319, 'C', 4321-4323, 'T', 4325-4327, 'A', 4329-4331, 'C', 4333-4335, 'G', 4337-4339, 'C', 4341-4343, 'T', 4345-4347, 'A', 4349-4351, 'C', 4353-4355, 'G', 4357-4359, 'C', 4361-4363, 'T', 4365-4367, 'A', 4369-4371, 'C', 4373-4375, 'G', 4377-4379, 'C', 4381-4383, 'T', 4385-4387, 'A', 4389-4391, 'C', 4393-4395, 'G', 4397-4399, 'C', 4401-4403, 'T', 4405-4407, 'A', 4409-4411,

Db 605 SVELNGNGKMKR-----SQSETKLNEPLKRWEEB-TRIKEARLR-----EEN 647
Qy 84 AQCEPAEPDAKSRITYVARNPEPTPVVYGEKPSKDPNTSEIRQSDVGDHRHP 143
Db 648 DRRERVAVEKAENKRLKAALQOEKPKIKNGEAREKAE----- 698
Qy 144 QKKKAGLGLKETTLLMOTLNTLTSTPEEKLAALCKKYAELLBEHRNSQKMLLOKQSQ 203
Db 699 QEKOM-----KE-----QQELEQLKAEFAKEHENRMRAPFALEQEKERR 739
Qy 204 L--VQEKD-----HLRGHSAVL-----ARSKLSLCRELQHRNLSKEGVQAR-----EEE 251
Db 740 INEAREKEENRRIKAREKAELEQRLKATLEQEKERQIKERQEREEENRERAKEVLEQA 799
Qy 252 EKKKEVTSHPQVTLNDIQLOMEQHNHNSKLRQENNELAERLKLTEQVELREEHIDKVP 311
Db 800 ENERKUEALEQENSRRLKTEKENGKKLREALELEKEKRLLEAFERAB--IERL 857
Qy 312 KHKDLOQLVDALQQA--QEMLKAEERHQ-----REKDFLLKAEVESQMC 358
Db 858 K-EDLSQEMRMRLQAKGRRLHRENQEHQENRKOHEVSGESDEKERDACEMEKTC 916
Qy 359 LMKQQ--ETHLKOQALYITEKFPONTLS-----KSSEVFTTFQEMKMTKKIKLEK- 411
Db 917 TTKEARHGEQSSNESLSTLEBENSIDNVSNKQKEEGECTRORESMSARTCPWKVPEKT 976
Qy 412 -----ETTMVRSWES-----SNKALLEMAEBKTVRDKEGLEQVXI 448
Db 977 LKDSQKEGTNEMDADTLPERNEETPRLGENGCGNQNGESGEESTSVENIIG----- 1031
Qy 449 ORLEKLCRALOTERN--DLNCRVQDLSAGGGSITD-----SGPERRPBGPAQAPSPRVT 503
Db 1032 GKLEQSKNSBTSKDSVLKRVGLKTEVEERLEDVVGVRDQNPPE-----ESKAPKTS 1087

RESULT 43
A23767
myosin heavy chain, fast skeletal muscle - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 05-Jun-1988 #sequence_revision 05-Jun-1988 #text_change 20-Jun-2000
C:Accession: A23767
R:Saez, L.; Leinwand, L.A.
Nucleic Acids Res 14, 2951-2969, 1986
A:Title: Characterization of diverse forms of myosin heavy chain expressed in adult human skeletal muscle
A:Reference number: A93616; MUID:86176778; PMID:241254
A:Accession: A23767
A:Molecule type: mRNA
A:Residues: 1-876 <SAE>
A:Cross-references: GB:X03740; NID:g34839; PIDN:CAA27380.1; PID:g1335218
A:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; muscle; skeletal muscle

Query Match 8.6%; Score 232; DB 2; Length 876;
Best Local Similarity 20.8%; Pred. No. 0.0069;
Matches 110; Conservative 88; Mismatches 168; Indels 162; Gaps 17;

Qy 38 KPEGQARTAQSGALRDVSELSQLELDILSTYCVNNQGGPGEDGAQGPAPEDAEKS 97
Db 34 KIEBQALGMO--LQKKIKELQAKIEL-----EEIEAERTSAKAKL 76
Qy 98 RTTYVARNGEPTPVVYGEKPSKDPNTSEI--RQSDVGDHRHPQEKKKAGLGKE 155
Db 77 RSDLSK-----ELEBISERLBEAVGATSTQIEMNKKRE-----AE 111
Qy 156 ITLLMOTLNTLTSTPEKLAALCKK-----YALLEEHNNSQKVKLLQKQSLVQEKDH 210
Db 112 PQMRDLDEEATLQHEATAATLKKHGHADSVAGELGEQIDNLRVKKLEKSEKMWELDD 171
Qy 211 LRGEHSAVLARKSLKESLCREIQ-----RHNR 238
Db 172 LASNNETWSKAGNLEKCKRALEDQLSEIKTKEEQRLINDLTAQRLAQNQVEYSRQ 231

Qy 239 LKE-----EGVQARBEERKEKVTSH-FQVTLNDIQLOMEQHNER 278
Db 232 LDKOTLTQLSRGKQAFPTQOIBELKQLEBEIKAKSALAHALQSSRHDCDLLREQYEE 291
Qy 279 -----NSKLR-----QENNELAERLKLTEQVELREEHIDKVPKH 313
Db 292 QEAKAELORAMSKANSEVAQWRTKYETDAIORTSELEBAKCKLAQRLQDAEEHV----- 345
Qy 314 KDLQOQLVDALQQAQEMLKAEERHQBKDFLLKAEVESQSMCELMQOETHLQKQAL 373
Db 346 -----EAVNAKAS-----LETKQRLQNEVEDLMIDVERTNAACAALDKQTNFKDLAE 396
Qy 374 YTEKEEPONTLSKSSVFTTFQEMKMTKKIKLEKETTMYRSRWNSSNKALLEMAEB 433
Db 397 WKQKCEE-----THAVLESFQKESRSLSTELFKI-----KNAYESLDQLETLKRE 442
Qy 434 KTVRDKELEGL--QV-----KIQRIEKLCRALQOTERNDLNKRVDLSA 474
Db 443 KNNLQQLQISDITQEAEGGKRIHELEKIKKQVEQKSELALEAEAA 490

RESULT 44
S06006
myosin beta heavy chain, cardiac muscle [similarity] - rat
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: S06006; S07536; I67441; A02989
R:Kraat, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.
Nucleic Acids Res 17, 7529-7530, 1989
A:Title: Complete nucleotide sequence of full length cDNA for rat beta cardiac myosin b
A:Reference number: S06006; MUID:90016823; PMID:2798112
A:Accession: S06006
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-1935 <KRA>
A:Cross-references: EMBL:X15939; NID:G56656; PIDN:CAA34065.1; PID:G56657
R:McNally, E.M.; Kraat, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.
J. Mol. Biol. 210, 665-671, 1989
A:Title: Full-length rat alpha and beta cardiac myosin heavy chain sequences. Comparison
A:Reference number: S07535; MUID:90333919; PMID:2614840
A:Accession: S07536
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-950, 'RK', 953-1935 <MCN>
R:Mahdavi, V.; Lompre, A.M.; Chambers, A.P.; Nadal-Ginard, B.
Eur. Heart J. 5, 181-191, 1984
A:Title: Cardiac myosin heavy chain isozymic transitions during development and under p
A:Reference number: I53305; MUID:85179510; PMID:6241892
A:Accession: I67441
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1871-1935 <RES>
A:Cross-references: GB:X32698; NID:G205598; PIDN:AAA41659.1; PID:G205599
R:Mahdavi, V.; Periasamy, M.; Nadal-Ginard, B.
Nature 297, 659-664, 1982
A:Title: Molecular characterization of two myosin heavy chain genes expressed in the ad
A:Reference number: A02988; MUID:82220036; PMID:7045682
A:Accession: A02989
A:Molecule type: mRNA
A:Residues: 1524-1528, 'V', 1530, 'R', 1532-1730, 'H', 1732-1783, 'K', 1785-1850, 'N', 1852-1857,
A:Cross-references: GB:J00752; NID:G205577; PIDN:AAA41654.1; PID:G205578
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; hydrolase; methylat
P:88-766/Domain: myosin motor domain homology <MYOT>
P:178-185/Region: nucleotide-binding motif A (P-loop)
P:548-585/Region: actin binding #status predicted
P:655-677/Region: actin binding #status predicted
P:839-1935/Domain: coiled coil #status predicted <COI>
P:839-1279/Region: 82
P:1280-1935/Region: light meromyosin
F:129/Modified site: N6,N6-trimethyllysine (lys) #status predicted
F:184/Binding site: ATP (Lys) #status predicted

F:695,705/Active site: Cys #status predicted

Query Match 8.6%; Score 232; DB 1; Length 1935;

Best Local Similarity 22.3%; Pred. No. 0.016;

Matches 116; Conservative 81; Mismatches 189; Indels 134; Gaps 16;

QY 25 VBAEGPGSCAPRXPBGQAARTAQSGALRDVSEELSRQLEDILSYCVDDNQGPGDGA 84

DB 1094 IDEQALGSQLQKLELQARI-----ELISEELE-----1123

QY 85 QGRPAEPDAKSRRTYVARGEPEPTVVYGGKPSKGDPTTEIROSDEVDGRDHRPPQ 144

DB 1124 -AERTARAKVEKRLSDLSRELEB-----ISRLBEAGGATSVQI-----EMKKREA 1169

QY 145 EKKKAGLGKELTLLWTLNTLSTPEKLAALCKV---AELLBEHNSOKOMKLLQKK 200

DB 1170 EFQKWRDLLEATL-----QHEATAALKKHADSVAELGQIDNLRVQKLEKE 1220

QY 201 QSQLVOEKHLRGESHKAVLARSKLESICR-----EL 232

DB 1221 KGEFKLEEDVTSNMEQIKAKANLERMCRITLEDQNEHRSKABEFTORSVNDLTRQAKL 1280

QY 233 QRHN-----RSIXE-----EGVORAREEKKKEVTSH-FQVTLNDI 268

DB 1281 OTENGELSOLDEKALISQLRGKLYTQQLDLKRLQLEEVKAKALAHALQARHDC 1340

QY 269 QLOMEQHNRNSK-----LQENWELAEKLLIEQYELREZHIDKVFHKDLQOLV 321

DB 1341 DLLREQYEBETEAKALORVLISKANSVAQMRKTYETDAIQRTSELSEA--KKKLAQRLQ 1398

QY 322 DAKLOQAQ-----MLKEAEERHQREKDFLLKEAVESQRMCELMKQOETHLKOQALY 374

DB 1399 DA--BEAVAVNAKSSLEKTKHRLQNEHDLVWDVSNAAAALDKQKQNFKILVEW 1456

QY 375 TKPFPQNTLSKSEVPTTFQENKEMTKIKKLEKTTMYRSRWSSNKALLEMAEK 434

DB 1457 KORYEESQELSSQKARSLSLTEL-----FKLNAYESLSHLETFKRNKNLQEEISDL 1512

QY 435 TVRDELEGLQVKIQKLEKLCALQTERNDLNKRQDLISA 474

DB 1513 T---EQLSGTGSIHLEKIRQLAEKLELQSALEAEA 1549

RESULT 45

A59236

embryonic muscle myosin heavy chain - sea squirt (Halocynthia roretzi)

C:Species: Halocynthia roretzi

C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 08-Sep-2000

C:Accession: A59236

R:Araki, I.

Submitted to GenBank, February 1999

A:Reference number: A59236

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1927 <ARA>

A:Cross-references: GB:D45163; NID:q1197167; PIDN:BA08111.1; PID:q1197168

A:Experimental source: clone lib lambda gtl1; dev stage tailbud embryo

C:Genetics:

A:Gene: MRCenb

C:Superfamily: myosin heavy chain; myosin motor domain homology

F:89-766/Domain: myosin motor domain homology <MWO>

Query Match 8.6%; Score 231.5; DB 2; Length 1927;

Best Local Similarity 21.0%; Pred. No. 0.016;

Matches 111; Conservative 96; Mismatches 196; Indels 125; Gaps 16;

QY 8 EAGPGGAERPSQAAPAV-----EAGPGSSQAAPKPEGAQARTAQSGALRDVSELS 60

DB 1363 KANAFAVQWRKRYETDAIQRTTELEAKKLLATLQAEQVEATQAKCASLQRTKRLQ 1422

QY 61 RQLRDLILSYCVDDNQGPGDGAQPAEPDAKSKRTYVARNGEPEPTFVYGGKPS 120

DB 1423 GELEDLTI-----DLERSNSAAAAL-----DKKQR 1447

QY 121 KGDPTTEIROSDEVDGRDHRPPQEKKAAGLGKEITLLM-----QTLNLTSTPEKLAAL 176

DB 1448 NFDKVLAKSKQEBE--EIOVLEQOAKKEARGLSLTFPKKNSYERSLDALETVKRKNQL 1505

QY 177 CKYVALLBEHNSOKOMKLLQKXQSQLVQEKDHLR-----GEHSKAVLARSKL 225

DB 1506 QSEIADLTQLQEGGKRSHELEKAKTLEHERNEIQAALEBAEAGISGESKVLRLQVEL 1565

QY 226 ESLCRLQHNRLSKBEGVQVRAEERKE-----VTSHFQVTLND 267

DB 1566 AQTKQDPER-RLSEKSEETENTRRNQRALESMTTLDSEKSRQEAVERMKCKMBGLND 1624

QY 268 IOLQMEQHNRNSKLRQENMELAEKLLIQYELREZHIDKVFHKDLQOO-----319

DB 1625 LSIQCHATROASEQSKSVKTFQAHVKDLLELDVDESQHS-----DLQSQPAVIERE 1678

QY 320 -LVDKLOQAQEMLKEAEERHQREKDFLLKEAVESQRMCELMKQOETHLKOQALYTKF 378

DB 1679 NLTKABIDELRSALQAE-----RGRKLAETELLESSESRNLLHTQNTALINQ-----KKKL 1730

QY 379 E-BFQNTLSKSEVPTTFQENKEMTKIKKLEKTTMYRSRWSSNKALLEM 430

DB 1731 EGELQNVQSEVEAVQENRABEAKKAITDAAMWAEELKEQDI-----SSHLEPKK 1784

QY 431 ABEKTVRDKB-----LEGQVKIQKLEKLCALQTERNDLNKR 468

DB 1785 NTEQTVKQLQRLDEAEQVALLGKGGKQVQKLETRVRELENELEDSQOR 1832

RESULT 46

S06005

myosin alpha heavy chain, cardiac muscle [similarity] - rat

N:Alternate names: alpha-myosin heavy chain

C:Species: Rattus norvegicus (Norway rat)

C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002

C:Accession: S06005; S07535; A20971; A02988; 153305

R:McNally, E.M.; Gianola, K.M.; Leinwand, L.A.

A:Title: Complete nucleotide sequence of full length cDNA for rat alpha cardiac myosin

A:Reference number: S06005; MUID:90016822; PMID:2798111

A:Accession: S06005

A:Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-1938 <MCN>

A:Cross-references: EMBL:X15938; NID:G56654; PIDN:CAA34064.1; PID:G56655

R:McNally, E.M.; Kraft, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.

J. Mol. Biol. 210, 663-671, 1989

A:Title: Full-length rat alpha and beta cardiac myosin heavy chain sequences. Comparison

A:Reference number: S07535; MUID:90133919; PMID:2614840

A:Accession: S07535

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1938 <MC2>

R:Mahdavi, V.; Chambers, A.P.; Nadal-Ginard, B.

Proc. Natl. Acad. Sci. U.S.A. 81, 2626-2630, 1984

A:Title: Cardiac alpha- and beta-myosin heavy chain genes are organized in tandem.

A:Reference number: A20971; MUID:84194059; PMID:6585819

A:Accession: A20971

A:Molecule type: protein

A:Residues: 1-12, 'Ap', 14-45, 'A', 47-50, 'Ap', 53-81, 'E', 83-86, 'Q', 88-109, 111-133, 'H', 135-1.

R:Mahdavi, V.; Periasamy, M.; Nadal-Ginard, B.

Nature 297, 659-664, 1992

A:Title: Molecular characterization of two myosin heavy chain genes expressed in the ad

A:Reference number: A02988; MUID:82220036; PMID:7045682

A:Accession: A02988

A:Molecule type: mRNA

A:Residues: 1512-1574, 'S', 1576-1721, 'N', 1723-1851, 'N', 1853-1869, 'N', 1871-1933, 'I', 1935-

A>Note: There are 10 or more myosin heavy chain genes in the rat, at least two of which

R:Mahdavi, V.; Lompre, A.M.; Chambers, A.P.; Nadal-Ginard, B.

Eur. Heart J. 5, 181-191, 1984

A:Title: Cardiac myosin heavy chain isozytic transitions during development and under pa
A:Reference number: 153305; MUID:85179510; PMID:6241892
A:Accession: 153305
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1872-1933; 'I', 1935-1938 <RES>
A:Cross-references: GB:M2697; NID:G205597; PIDN:AAA41658.1; PID:G205597
A:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; hydrolase; methylate
P:87-767/Domain: myosin motor domain homology <MMOT>
P:177-184/Region: nucleotide-binding motif A (P-loop)
P:1548-585/Region: actin binding #status predicted
P:1658-678/Region: actin binding #status predicted
P:1840-1938/Domain: coiled coil #status predicted <COI>
P:1840-1938/Region: S2
P:1281-1938/Region: light meromyosin
F:128/Modified site: N6,N6-trimethyllysine (lys) #status predicted
F:183/Binding site: ATP (lys) #status predicted
F:696,706/Active site: Cys #status predicted

Query Match 8.6%; Score 231.5; DB 1; Length 1938;
Best Local Similarity 24.4%; Pred. No. 0.016;
Matches 146; Conservative 80; Mismatches 197; Indels 175; Gaps 25;
QY 12 EGAORPQAPAVAEAGSGSQAPRPEGAQARTAGSALRDVSELSQLEDLSTYC 71
DB 1036 EGSLEQEKVMDLE-----RAKRLGDLKLTQES--IMDLEND-KLQLEKLLKKE 1085
QY 72 VDNQGGPBGAGQCEPAPEDA-----EKSRTYVAR-----NGEPPTPVVYGEKEP 119
DB 1086 FDISQ-----QNSKTEDEQALALQKLEKQENQARIEELEEAEARTAKVEKLR 1137
QY 120 SKGDPNTEIRQS-DEVG-----DRHRRPOEKKKAGLKGKEITLLMOTLTLSTPEE 171
DB 1138 SDLTLESEISERLEEAGGATSVQIEMNKKREAFQKRRDLEATL-----QHEA 1188
QY 172 KLAALCKKY-----AELLLEHNSOKMKLQKQSQLVQSKHURGHSHKAVLARSKLES 227
DB 1189 TAAALRKKHADSVAGLGEQIDNLRQVKQLEKSEKSEFKLEDDVTSNMEQIIKAKANLEK 1248
QY 228 LCR-----ELORHNSLKEEGVQAR-----KLEKETTTMYR 248
DB 1249 VSRILEDOANRYRVKLESAQSLNDFTQRAKLQTEGELARQLEKEKALIQLTGKLS 1308
QY 249 -----ESEEKKEVTSH-FQVTLNDIQLOQHEQNER-----NSK 281
DB 1309 YTOQMEDLKROLEBEGKAKNALAHQSSRHDCLLSEQYEEEMEAQELQVLSKANSE 1368
QY 282 LR-----QENMELAEELKLEIYELREEHIDKV-----FKHK---DLQ 318
DB 1369 VAQWRTKYETDAIQRTTELEBAKKLAQRLQDAEAEVAVNAKSSLEKTKHRLQNEIED 1428
QY 319 QLVDAKQQAQEMKLEBAERHQRKDFLLKBAVESQRMCELMKQOETHLQKQALALYTEKF 378
DB 1429 LMVDVERSNA---AAAALDKQRNFDKILAE--WKQKYESSQSELESSQKARSLSLSELF 1483
QY 379 EEFQNTLSKSSEVFTTFKQEMKMTKKI-----KLEKETTTMYR 417
DB 1484 -KLKVAZESLEHETFGKNNKQESISDLTEQLGSGGNVHELEKIRQLEVEKLEIQ 1542
QY 418 SRWSSNKALLEMAEKTVRQKLEGLQVKIQRLKLCJ---RALQTERNDLKNRVOD 471
DB 1543 SALEEA-BASLEHEEGKILR-AQLEFNQIKAEIRKLEKADBEQAQRNEL--RVVD 1596

RESULT 47
149464
alpha cardiac myosin heavy chain - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence revision 02-Jul-1996 #text change 02-Feb-2001
C:Accession: 149464; 149463; 149462; 149461; 149604
R:Quinn-Laguer, B.K.; Kennedy, J.E.; Wei, S.J.; Beisel, K.W.
Genomics 13, 176-188, 1992

A:Title: Characterization of the allelic differences in the mouse cardiac alpha-myosin
A:Reference number: A38207; MUID:92250040; PMID:1577481
A:Accession: 149464
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1938 <RES>
A:Cross-references: GB:M76601; NID:G191623; PIDN:AAA37162.1; PID:G191624
A:Accession: 149463
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-193, 'D', 195-837, 'S', 839-955, 'N', 957-1938 <RE2>
A:Cross-references: GB:M76600; NID:G191621; PIDN:AAA37161.1; PID:G191622
A:Accession: 149462
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1938 <RE3>
A:Cross-references: GB:M76599; NID:G191619; PIDN:AAA37160.1; PID:G191620
A:Accession: 149461
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-544, 'A', 545-1938 <RE4>
A:Cross-references: GB:M76598; NID:G191617; PIDN:AAA37159.1; PID:G191618
R:Gulick, J.; Subramanian, A.; Neumann, J.; Robbins, J.
J. Biol. Chem. 266, 9180-9185, 1991
A:Title: Isolation and characterization of the mouse cardiac myosin heavy chain genes.
A:Reference number: 149604; MUID:91225025; PMID:2026617
A:Accession: 149604
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-67 <RES>
A:Cross-references: GB:M62404; NID:G192609; PIDN:AAA37424.1; PID:G192610
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: ATP; cardiac muscle; heart; nucleotide binding; P-loop
F:88-768/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 8.6%; Score 231.5; DB 2; Length 1938;
Best Local Similarity 24.2%; Pred. No. 0.016;
Matches 143; Conservative 81; Mismatches 195; Indels 173; Gaps 24;
QY 12 EGAORPQAPAVAEAGSGSQAPRPEGAQARTAGSALRDVSELSQLEDLSTYC 71
DB 1037 EGSLEQEKVMDLE-----RAKRLGDLKLTQES--IMDLEND-KLQLEKLLKKE 1086
QY 72 VDNQGGPBGAGQCEPAPEDA-----EKSRTYVAR-----NGEPPTPVVYGEKEP 119
DB 1087 FDISQ-----QNSKTEDEQALALQKLEKQENQARIEELEEAEARTAKVEKLR 1138
QY 120 SKGDPNTEIRQS-DEVG-----DRHRRPOEKKKAGLKGKEITLLMOTLTLSTPEE 171
DB 1139 SDLTLESEISERLEEAGGATSVQIEMNKKREAFQKRRDLEATL-----QHEA 1189
QY 172 KLAALCKKY-----AELLLEHNSOKMKLQKQSQLVQSKHURGHSHKAVLARSKLES 227
DB 1190 TAAALRKKHADSVAGLGEQIDNLRQVKQLEKSEKSEFKLEDDVTSNMEQIIKAKANLEK 1249
QY 228 LCR-----ELORHNSLKEEGVQAR-----KLEKETTTMYR 248
DB 1250 VSRILEDOANRYRVKLESAQSLNDFTQRAKLQTEGELARQLEKEKALISQLTGKLS 1309
QY 249 -----ESEEKKEVTSH-FQVTLNDIQLOQHEQNER-----NSK 281
DB 1310 YTOQMEDLKROLEBEGKAKNALAHQSSRHDCLLSEQYEEEMEAQELQVLSKANSE 1369
QY 282 LR-----QENMELAEELKLEIYELREEHIDKV-----FKHK---DLQ 318
DB 1370 VAQWRTKYETDAIQRTTELEBAKKLAQRLQDAEAEVAVNAKSSLEKTKHRLQNEIED 1429
QY 319 QLVDAKQQAQEMKLEBAERHQRKDFLLKBAVESQRMCELMKQOETHLQKQALALYTEKF 378
DB 1430 LMVDVERSNA---AAAALDKQRNFDKILAE--WKQKYESSQSELESSQKARSLSLSELF 1484
QY 379 EEFQNTLSKSSEVFTTFKQEMKMTKKI-----KLEKETTTMYR 417

Db 1485 -LKRAYESLEHETFRKRNKXQBSISDTEQLEGQKQVHELEKIRKQLEVKLELQ 1543
QY 418 SRWESSNALLEMAEKTVRKELGQWKIQRLEKLC-...RALQTERNDL 465
Db 1544 SALEEA-BASLEHSGKILR-AQLSFNQIKABIERKLAEDMEQAKRNEL 1593
RESULT 48
S74245
serine/threonine-specific protein kinase (BC 2.7.1.1-) isoform II, Rho-associated - mouse
C:Species: Mus musculus (house mouse)
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 24-Sep-1999
C:Accession: S74245
R:Nakagawa, O.; Fujisawa, K.; Ishizaki, T.; Saito, Y.; Nakao, K.; Narumiya, S.
PES Lett 392, 189-193, 1996
A:Title: ROCK-I and ROCK-II, two isoforms of Rho-associated coiled-coil forming protein
F:90-354/Domain: protein kinase homology <kin>
A:Reference number: S74244; MUID:96368048; PMID:8772201
A:Accession: S74245
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1388 <NAK>
A:Cross-references: EMBL:U58513; NID:g1514697; PIDN:AAC53133.1; PID:g1514698
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase C zinc
C:Keywords: phosphotransferase; serine/threonine-specific protein kinase
F:1261-1315/Domain: protein kinase C zinc-binding repeat homology <KZN>
Query Match 8.5%; Score 231; DB 2; Length 1388;
Best Local Similarity 21.5%; Pred. No. 0.012;
Matches 133; Conservative 84; Mismatches 172; Indels 230; Gaps 23;
QY 91 PEDAKSR-...TYVARNGEPE-...PTPVV-... 113
Db 316 PDETBISKXKGLICAPLTDREVLGRNGVEIKQHPFFKNDQNWMDNIRETAAPVPEL 375
QY 114 -...YGEKEPKSGDPNTEI-...ROSDEVG 136
Db 376 SSDIDSNFDDIEDDKGDEVEFPKAFVGNQLPIGTYFERNLLSDSPPCRENDIQ 435
QY 137 DRDHRRPQ-...KKAKGLGKEITLMTL 163
Db 436 TRKSESQEIQKGVALBEHLSSEVOAKELEQCKKSINTELEKTAKELEEBEITLRKSE 495
QY 164 NTLSTPPEKLAALCKKYAEL-...LEEHRNSOK-QMKLLQKK-... 200
Db 496 STQLREKALKQHNAYQKADHEADKKNLENDVNSLKOLEDLAKKNQSSQISTE 555
QY 201 -...QSQLVQKQHLRGEHSAV-LARSKLES-LCRELQHRNSLKER-... 242
Db 556 KYNQLQKQLEHANAALLRTESPTAELRKTQAESKQIQLESNNRDLQDNKCLLETAKLK 615
QY 243 -...GVQARAEERKEKVTSHFPQVTLNDIQ-...LQMEQH 275
Db 616 LEKEFINLSALESERRD-...TKGSEIINDLQGRISGLEBDLKTGKALLAKVELEKRL 672
QY 276 NERNKLRQ-...NMLEARLKKLBOYHLRREHDKVFKHLOQQLVD-...-A 323
Db 673 QEKLTDLKEKKNMBIDMTYQLVKVIQQLSLEFEAS-...HKTTRALRADKRVKIVESTEA 728
QY 324 KLCQQAQEMLKE-ABERHOREK-DFLLKKAIVE-...SQRMELMKQOETHLKQ 369
Db 729 KSEAMKEMEKLLERSLKQKVENLLAEKRCISILDCLAQSQKLNELLKQDV-LNE 787
QY 370 QLAALYTERKEE-...FNTLSKSEVFTT-...FKQEMEKTKKIKKLEKETT 415
Db 788 DVYRNTLTKEQTKRCLMNDLKNQMTQVNTLKMSEXQIKQENHLMEMHONLEKONTE 847
QY 416 YRSRWESSNALLEMAEE-...KTVRDEKESLQVKIQRLEKLCRALQTERNDLNRK 468
Db 848 LKREQDADGQMKELQDLQEAQYFSTLYKTQVRELKEENBEKTKLCKELQKQKQDLQDE 907
QY 469 VQDLSAGQCGSLTDSGPER 487

Db 908 RDSLAQAQLEITLTKADSEQ 926
RESULT 49
T08880
NMDA receptor-binding protein yotiao - human
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08880
R:Lin, J.W.; Wyszynski, M.; Madhavan, R.; Sealock, R.; Kim, J.U.; Sheng, M.
J. Neurosci. 18, 2017-2027, 1998
A:Title: Yotiao, a novel protein of neuromuscular junction and brain that interacts wit
A:Reference number: Z16511; MUID:98151389; PMID:9482789
A:Accession: T08880
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1642 <LIN>
A:Cross-references: EMBL:AF026245; NID:g2623067; PIDN:AAB86384.1; PID:g2623068
C:Genetics:
A:Map position: 7q21-22
C:Keywords: brain; cerebral cortex; coiled coil; neuromuscular junction; skeletal muscul
Query Match 8.5%; Score 230.5; DB 2; Length 1642;
Best Local Similarity 18.4%; Pred. No. 0.015;
Matches 119; Conservative 110; Mismatches 195; Indels 223; Gaps 23;
QY 31 GSSQAPRPEGAQARTAGSALRDVSELSROLEDILSTYCVDNQ-...GGPGEQ 83
Db 161 GAQDSPTHELMWSELA--GKQHEI-ELNARELSEMRVTYGTGLQLOLQEFEEAIAKORDG 217
QY 84 -...AQGEPAPEDEAKSRTYVARNGEPEPTVYVYGEKPSKGDPT-... 126
Db 218 ITTUTANLQARRKEDTMEFELETLQSKLQIQFOLOQASTLRNSTHSTSTAADLLQ 277
QY 127 -...EBIROSDVGDHRRPQEKRAKGLGKEITLMTLNTLSTPBEKLAALCK-... 179
Db 278 AKQILTHQOOLEEQHLLDYQKKEDFTQISFLQEKIKVYEMEQDKKVENSKZBIQ 337
QY 180 -...YAELESEHNSOKQ-...KLQKQSQLVQKQHLRGEHSAVLA 223
Db 338 EKETIIEBLNTKIIEBKRTLELKDGLTADKLGELOQIVQKQEK--NMKLELTNS 395
QY 224 K-...LESCLRELQHRN-...SLKBEVGQARABEEBKRKE-... 256
Db 396 KQKERQSSEIKQLMGTVEELQKRNHDKSQFETDIVORMEQETORLEQLEAELEDMYQ 455
QY 257 -...-VTSH-...FOVTLNDIQLQME 273
Db 456 QIVQKQELIRQHQMAQEMKTRHGENNALSYSNITWNEDQIKLMNVAINSLIKLQ 515
QY 274 QHNRNSKL-...QNNMELABRLKGLIEQYELREHIDKVK-...-HX 314
Db 516 DTNSQKELKEELGILIEKCALQRLQEDLVBELSFREQIQRARQTIABQESKLENAHK 575
QY 315 DLQO-QLVDAKLQQAQEMLKEAERH-...-OREKDPELLKEAVESQRM-... 356
Db 576 SLSTVEDLKAEIVSASERKELELKHAEVNTYKLEMLEKKNVLDLDMASQEALE 635
QY 357 -...CELKKQO-...THLKQOL-...ALYTEKF 378
Db 636 RLRTQLLPSEBELSKLEBDLEIEHRINIEKLNGLGHYKQIQDGLQENMSQKLETMQF 695
QY 379 BEPQNTLSKSEVFT-...TFKQEMEKTKKIKKLEKETT 427
Db 696 EK-DNLITKQNLILIEISKDLQOOSLVNSKSEKTIQINELQKEILR-... 744
QY 428 LEMAEKTVRKELGQWKIQRLEKLCRALQTERNDLNRKVDLSA 474
Db 745 -QEEKKGTLEQVQELQKLTLEK-...QMKENDLQEKFAQLEA 787
RESULT 50

```

Query Match      8.5%; Score 230.5; DB 2; Length 1939;
Best Local Similarity 24.4%; Pred. No. 0.018;
Matches 146; Conservative 80; Mismatches 197; Indels 175; Gaps 25;

QY 12 EGAQRPQAAPAVAPAEAGPSSQAPRPEGAQAARTAQSGALRDVSEELSDILSTYC 71
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 1037 EGSLEQEKVRMDLE-----RAKRLLEGDLNVTCES--IMDLND-KLQLEKXKKE 1086

QY 72 VDNVQGGFGDGAQGEAPABEDA-----EKSTYVAR-----NGEPTPVVYGEKEP 119
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 1087 PDISO-----QNSKIEQALALQKQLKENQARIELEEBLEAERTAKAVKRLR 1138

QY 120 SKGDPNTEFIROS--DEVG-----DRHRRPORKKAKGLGKITLLMQTLNLTSTPEE 171
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 1139 SDLTRELESEIERLEKAGGATSVQIENMKKREAFQKMRDRLEETL-----QHEA 1189

QY 172 KLAALCKY----KELLEEHNSQKMKLLQKQSQLVQEKDHLRGESKAVLARSKLES 227
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 1190 TAPALRKHAUSVAILGEQIDNQRVKQLEKSEFKLELDVDTNSNMEQIKKAKANLEK 1249

QY 228 LCR-----ELQRNRSUKBEGVQAR----- 248
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 1250 VSRTLQDQANRYRVKLBESQSRSLNDFDTQRAKLTQENGELARQLEBEGALISQULTROKLS 1309

QY 249-----ESEEKRRKEVTSH--FOVTLNDIQLQMEQHNR-----NSK 281
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 1310 YTOQMEDLQRLQEBEGKAKNALAHQSAHDCOLLREQYEEBMEAKAELQVLVLSKANSE 1369

QY 282 LR-----QENMELARIRKLLIEQVELAEEHIDKV-----FTHK---DLQQ 318
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 1370 VAQWTKYETDAIORTBELEBAKXKGLAARLODAEAVAVNAKCSSELEKTKHRLQNHIED 1429

QY 319 QLVADAKLQAQOMLKEAEERHOREKDFLLKAEAVESQRMCELMKQOFTHLKQOLALYTKFP 378
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 1430 LMVDVVERNA--AAALDKQARNDFDKILAE--WKQKYEESQSESSQKEARSILSTELF 1484

QY 379 EFPQNTLSKSSEVETTPQWMEKQTKI-----KKLEKETTMYR 417
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 1485 -KKKNAYEESLHEUETTFKRNKQNLQESIDLTQLEGGKNVHELEKVRQKVEKQEMQLQ 1543

QY 418 SRWESSNKALLMEAEKTVRDKELEGVQKIQBLEKLC-----RALQTERNDINLRQVD 471
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 1544 SALEBA-EASELEHEGKILR-AOLEFNQIKAEITERKLAEKDEMEQAKNELL--RVVD 1597

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 12:41:25 ; Search time 27 Seconds
(without alignments)

1022.118 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 2702

Sequence: 1 KSSPQPEAGPEGQERPSQ.....APSTASGQTGPQPTSARA 530

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : SwissProt 42:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1315	48.7	259	YL14_HUMAN	P40222 homo sapien
2	264.5	8.8	1084	MYSS_RABIT	P02562 oryctolagus
3	261.5	9.7	1407	TRHY_RABIT	P27709 oryctolagus
4	259	9.6	886	RA50_ARCFU	Q92230 archaeoglob
5	259	9.6	1938	MYSS_CHICK	P13538 gallus gall
6	256	9.5	2871	DESP_HUMAN	P15924 homo sapien
7	252.5	9.3	880	RA50_PYRAB	Q9uzc8 pyrococcus
8	251.5	9.3	1549	TRHY_SHEEP	P22793 ovis aries
9	251.5	9.3	1690	CI90_DROME	Q9vjf5 drosophila
10	249	9.2	1938	MYH4_RABIT	Q28641 oryctolagus
11	246	9.1	1427	REST_HUMAN	P30622 homo sapien
12	245	9.1	1937	MYH8_HUMAN	P13535 homo sapien
13	245	9.1	1940	MYH3_CHICK	P02565 gallus gall
14	244.5	9.0	1410	RRB1_HUMAN	Q9p2e9 homo sapien
15	243.5	9.0	993	SCP1_MOUSE	O62209 mus musculus
16	242.5	9.0	4473	PLE1_CRIGR	O9j155 cricetus
17	242	9.0	1534	RRB1_CANPA	Q28238 canis fam
18	242	9.0	3259	GOB1_HUMAN	Q14789 homo sapien
19	240	8.9	2663	CENE_HUMAN	Q22244 homo sapien
20	239.5	8.9	1898	TRHY_HUMAN	Q07283 homo sapien
21	239.5	8.9	1939	MYH1_HUMAN	P12882 homo sapien
22	239	8.8	4684	PLE1_HUMAN	Q15149 homo sapien
23	238.5	8.8	4687	PLE1_RAT	P30427 rattus norv
24	237	8.8	1934	MYH7_MESAU	P13540 mesocricetu
25	237	8.8	1938	MYH2_HUMAN	Q9ukx3 homo sapien
26	236.5	8.8	1939	MYH6_HUMAN	P13533 homo sapien
27	236	8.7	978	RA50_AQUAR	Q07124 aquifex aeo
28	236	8.7	1411	EEA1_HUMAN	Q15075 homo sapien
29	234	8.7	1935	MYH7_HUMAN	P12883 homo sapien
30	234	8.7	1935	MYH7_PIG	P79293 sus scrofa
31	233.5	8.6	845	SCP1_MESAU	O50561 mesocricetu
32	233	8.6	1941	MYH2_HUMAN	Q9ukx2 homo sapien
33	232	8.6	1935	MYH7_RAT	P02564 rattus norv

34	231.5	8.6	882	RA50_PYRFU	P58301 pyrococcus
35	231.5	8.6	1938	MYH6_MOUSE	Q02566 mus musculus
36	231.5	8.6	1938	MYH6_RAT	P02563 rattus norv
37	231	8.5	1605	RRB1_MOUSE	Q99p15 mus musculus
38	230.5	8.5	1939	MYH6_MESAU	P13539 mesocricetu
39	230.5	8.5	1939	AKA9_HUMAN	Q99996 h a-kinase
40	230	8.5	1939	MYH4_HUMAN	Q9v623 homo sapien
41	229.5	8.5	790	BEA1_MOUSE	Q8b166 mus musculus
42	229.5	8.5	2230	GOA1_HUMAN	Q13439 homo sapien
43	229.5	8.5	2238	GOA4_MOUSE	P91w5 mus musculus
44	229	8.5	1976	MYH4_HUMAN	P35580 homo sapien
45	229	8.5	2779	LVA_DROME	Q8ms1 drosophila
46	228	8.4	1938	MY5_AEQIR	P24733 aequipecten
47	228	8.4	2442	CBP2_HUMAN	Q9bv73 homo sapien
48	226.5	8.4	1790	USO1_YEAST	P25386 saccharomyc
49	226.5	8.4	1940	MYH3_HUMAN	P11055 homo sapien
50	226.5	8.4	2210	CENF_HUMAN	P14954 homo sapien
51	225.5	8.3	1940	MYH3_RAT	P12847 rattus norv
52	225	8.3	1919	INCE_HUMAN	Q9nq57 homo sapien
53	224	8.3	1961	MYH9_RAT	Q62812 rattus norv
54	224	8.3	1976	MYHA_BOVIN	Q27991 bos taurus
55	224	8.3	1976	MYHA_RAT	Q97961 vulpes vulp
56	223.5	8.3	1679	GCC2_MOUSE	P12270 homo sapien
57	223	8.3	492	M5_STEPS	O9x1x1 thermotoga
58	221.5	8.2	997	SCP1_RAT	Q91mh9 mus musculus
59	221.5	8.2	1102	MYSC_CHICK	P02569 rattus norv
60	221	8.2	1935	MYSS_CYPCA	P35418 taenia soli
61	220.5	8.2	2017	MYSN_DROME	Q90339 cyprinus ca
62	220	8.1	1330	KTN1_VULVU	Q99323 drosophila
63	220	8.1	2349	TPR_HUMAN	Q97961 vulpes vulp
64	219.5	8.1	852	RA50_THEMEA	P12270 homo sapien
65	219.5	8.1	2035	MY8A_MOUSE	O91mh9 mus musculus
66	219	8.1	1846	MY5B_RAT	P02569 rattus norv
67	218.5	8.1	863	MYSP_TAESO	P35418 taenia soli
68	218.5	8.1	1197	CING_HUMAN	Q9p2m7 homo sapien
69	218.5	8.1	1742	MY5C_HUMAN	Q9nqk4 homo sapien
70	218	8.1	1364	KTN1_CHICK	Q90631 gallus gall
71	218	8.1	1433	REST_CHICK	O42184 gallus gall
72	218	8.1	3321	PCN2_HUMAN	O95613 homo sapien
73	217.5	8.0	967	KINH_LOLPE	P16113 loligo peal
74	217.5	8.0	1972	MYHB_RABIT	P35748 oryctolagus
75	217.5	8.0	1978	MYHB_CHICK	P10587 gallus gall
76	216	8.0	757	P1BF_HUMAN	Q8wxw3 homo sapien
77	216	8.0	1191	CING_MOUSE	P59242 mus musculus
78	216	8.0	1357	KTN1_HUMAN	Q86up2 homo sapien
79	216	8.0	1583	GCC2_HUMAN	Q81wj2 homo sapien
80	215.5	8.0	729	GOA5_MOUSE	Q9gye6 mus musculus
81	215.5	8.0	866	MYSP_SCHJA	Q95870 schistosoma
82	215.5	8.0	1360	CING_XENLA	Q9p4d7 xenopus lae
83	215.5	8.0	1959	MYH9_CHICK	P4105 gallus gall
84	215.5	8.0	1960	MYH9_HUMAN	P35579 homo sapien
85	215.5	8.0	1972	MYH8_HUMAN	P35749 homo sapien
86	215	8.0	697	MPF1_LYCES	P93203 lycopersico
87	214.5	7.9	1721	ITN1_HUMAN	Q15811 homo sapien
88	214.5	7.9	2033	EVPL_HUMAN	Q92817 homo sapien
89	214.5	7.9	2245	MY5J_DICDI	P54697 dictyosteli
90	214	7.9	1286	SMC4_MOUSE	Q8G947 mus musculus
91	214	7.9	1957	SPOF_SCHPO	Q10411 schistosom
92	213.5	7.9	767	GOA1_HUMAN	Q92805 homo sapien
93	213.5	7.9	880	INCE_MOUSE	Q9w62 mus musculus
94	213.5	7.9	1714	ITN1_MOUSE	Q920r4 mus musculus
95	213	7.9	862	CSP2_MACFA	Q92852 macaca fasc
96	213	7.9	1875	MLP1_YEAST	Q02455 saccharomyc
97	213	7.9	2054	MY8A_HUMAN	Q92614 homo sapien
98	212	7.8	1300	DYNA_NEUCR	Q01397 neurospora
99	211.5	7.8	1027	KINH_MOUSE	P33175 mus musculus
100	211.5	7.8	2116	MY52_DICDI	P08799 dictyosteli
101	211	7.8	1224	DYNA_CHICK	P35458 gallus gall
102	211	7.8	1290	SMC4_XENLA	P50532 xenopus lae
103	210.5	7.8	863	MY5P_ECHGR	P35417 echinococcu
104	210.5	7.8	1597	CTRO_MOUSE	P49025 mus musculus
105	209.5	7.8	1044	YAF3_SCHPO	Q09857 schizosacch
106	209.5	7.8	2611	BP1E_MOUSE	Q912u8 mus musculus

FT	DOMAIN	455	1084	COILED COIL (POTENTIAL).
FT	VARIANT	405	L -> V.	
FT	VARIANT	408	V -> L.	
FT	VARIANT	421	E -> D.	
FT	VARIANT	423	S -> G.	
FT	VARIANT	426	K -> R.	
SQ	SEQUENCE	1084 AA;	125488 MW; 229CFD69A61E7F0 CRC64;	

Query Match 9.8%; Score 264.5; DB 1; Length 1084;
Best Local Similarity 22.7%; Pred. No. 0.00071;
Matches 131; Conservative 85; Mismatches 194; Indels 167; Gaps 20;

QY	12	EGAQRPSQAAPVAEAGFGSSQAPRKPEGAQARTAQSGALRDVSELSRQLEDILSTYC	71
DB	193	EGSLEQEKIRMDLE-----RAKKLEB-DLKLAEQSM-DIEND-KQOLDEKLKLE	242
QY	72	VDNQGQPED-----GAQGPAPEDAEKSTFYVARGNEPEPTVVVG	115
DB	243	FTNLQSKTEQALMTNLRISLEBELEASRAKAKORSIDLSRELE-----IS	296
QY	116	EKEPSKGDENTIEIRQSDVGDHRRPQEKKAKGLGKEITLLMQTLNLTSTPEBKLA	175
DB	297	ERLEAGGATSAQI-----EMNKKGEAEFEKMRDLEATL-----QHEATAAA	340
QY	176	LCKKY-----AELLEHRNSQKQKLLQKQSQQLVQEKHLRGHSHKAVLARSKLESCRE	231
DB	341	LRKKHADSVAEELGEQIDNLRVQKLEKEKSLKMEIDDLGNMVTVSAAKGNLEKQVCR	400
QY	232	L-----QRHNRSLKE-----	241
DB	401	LEQQLSEVTKKEEHOELNELNSAQARLHTESGFSRQLDEKDMVSQLSGGQFTQQ	460
QY	242	-EGVQARBEKKEVTSH-FQVTLNDILOMEQVNER-----NSKLR--	283
DB	461	IEGLKQLBEETKAKSALAHALQSSRRDCDLLEQYEEQEAQAEQAKELQRAMSKANSEVSQ	520
QY	284	-----QENNELAERLKLLEQVLEFEEHDKVFKHKDLQQLQVLDLQQAQOEMLKE	334
DB	521	RYKCTDALQRTHELEAKKLAQLQDAEHEV-----EAVNSKCA--LEK	565
QY	335	AEERHQENDFLKAEAVESQRMCELMQOETHLQQLALYKTEKFEFQNTL-----S	386
DB	566	TKQLQNEABDLMI DVERSNATCARMDKQKRNFDKVAEWKHKYEEQAELEASQESRS	625
QY	387	KSEVETTFK-----QEMKMTKKIKLEKTTMYRSWESSKALLWABEKTVDKE	440
DB	626	LSTEVFKVKNAYEESLDHLETLKRNKNJQELSDLTEQALASAKHILEKVKQKIDQE	685
QY	441	LEGLOVKIQRLB-----KLCLALQTERNDLKNKV	469
DB	686	KSELQALAEAGSLSEHSGKILR-IQLHLNQVKSSEI	721

RESULT 3
TRY RABIT STANDARD; PRT; 1407 AA.
AC P37709;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Trichohyalin.
GN THH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Fietz M.J., Rogers G.B.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: Intermediate filament-associated protein that associates
CC in regular arrays with keratin intermediate filaments (KIF) of the
CC inner root sheath cells of the hair follicle and the granular

QY	452	EKLCLALQTERNDLKNKVQSLDSAGGQSLTDSGERRPEGGAQAPSSPRVTEAPCPYGA	511
DB	181	EKLCLALQTERNDLKNKVQSLDSAGGQSLTDSGERRPEGGAQAPSSPRVTEAPCPYGA	240
QY	512	PSTEASQGTGQPEPTSARA	530
DB	241	PSTEASQGTGQPEPTSARA	259

RESULT 2
MYSS RABIT STANDARD; PRT; 1084 AA.
AC P02562;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, skeletal muscle (Fragments).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE OF 1-258.
RA Capony J.-P., Eizinga M.;
RT "The amino acid sequence of A 34,000 dalton fragment from S-2 of
RT myosin.";
RL Biophys. J. 33:148A-148A (1981).
RN [2]
RP SEQUENCE OF 259-428.
RX MEDLINE=85131142; PubMed=3972832;
RA Lu R.C., Wong A.;
RT "The amino acid sequence and stability predictions of the hinge
RT region in myosin subfragment 2.";
RL J. Biol. Chem. 260:3456-3461 (1985).
RN [3]
RP SEQUENCE OF 409-1084 FROM N.A.
RX MEDLINE=87304245; PubMed=3305014;
RA Maeda K., Szakiel G., Wittinghofer A.;
RT "Characterization of cDNA coding for the complete light meromyosin
RT portion of a rabbit fast skeletal muscle myosin heavy chain.";
RL Eur. J. Biochem. 167:97-102 (1987).
CC -I- FUNCTION: Muscle contraction.
CC -I- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC and 2 regulatory light chain subunits (MLC-2).
CC -I- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -I- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -I- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC meromyosin (MM) and 1 heavy meromyosin (HMM). It can later be
CC split further into 2 globular subfragments (S1) and 1 rod-shaped
CC subfragment (S2).
CC -I- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X05958; CAA29391.1; -;
CC FIR; A02985; A02985.
CC FIR; A05280; A05280.
CC FIR; S00084; S00084.
CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding; Multigene family.
CC NON TER 1 1
CC DOMAIN <1 >258 ALPHA-HELICAL TAILPIECE (S2).
CC NON CONS 258 259
CC DOMAIN <259 1084 RODLIKE TAIL (S2 AND LAM DOMAINS).
CC FT

layer of the epidermis. It later becomes cross-linked to KIF by isodipeptide bonds. It may serve as scaffold protein, together with involucrin, in the organization of the cell envelope or even anchor the cell envelope to the KIF network. It may be involved in its own calcium-dependent postsynthetic processing during terminal differentiation.

CC -1- SUBUNIT: Homodimer (Probable).

CC -1- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as the inner root sheath (IRS) of hair follicles and medulla, and in the filiform papillae of dorsal tongue epithelium (Probable).

CC -1- DEVELOPMENTAL STAGE: Expressed during late differentiation of the epidermis.

CC -1- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand calcium-binding domains. Domains 2-4, 6, and 8 are almost entirely alpha-helical, configured as a series of peptide repeats of varying regularity, and are thought to form a single-stranded alpha-helical rod stabilized by ionic interactions. Domain 5 is the most regular and may bind KIF directly by ionic interactions. Domains 5 and 7 are less well organized and may induce folds in the molecule. Domain 9 contains the C-terminus, conserved among different species.

CC -1- PTM: Substrate of transglutaminase. Some 200 arginines are probably converted to citrullines by peptidylarginine deiminase.

CC -1- SIMILARITY: In the N-terminal section; belongs to the S-100 family.

CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; Z13092; CA79519.1; --

CC PIR; S28589; S28589.

CC HSSP; P02633; 4ICB.

CC InterPro; IPR001751; CABP_S100.

CC InterPro; IPR002048; EF-hand.

CC Pfam; PF00036; ehand; 1.

CC Pfam; PF01023; S_100; 1.

CC ProDom; PD003407; CABP_S100; 1.

CC PROSITE; PS00018; EF_HAND; 1.

CC PROSITE; PS00303; S100_CBP; 1.

CC Keratinization; Calcium-binding; Repeat; Citrullination.

CC DOMAIN 1 91 S-100 LIKE.

CC CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).

CC CA_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).

CC SEQUENCE 1407 AA; 183781 MW; AE17D2A1S9P12B7F CRC64;

Query Match 9.78; Score 261.5; DB 1; Length 1407;

Best Local Similarity 23.5%; Pred. No. 0.0012;

Matches 117; Conservative 110; Mismatches 168; Indels 103; Gaps 22;

QY 26 EAEGFGSS-----QAPKPEGAQTAQSGALRDVSELSQLELDILSTCYVDNNGG 78

DB 514 QEQPGQWQWQEQEQAQRHTYAPGQSQURE-EELQR-----EKRRQE 561

QY 79 PGEDGAGEPAEPDAEKSRITYVANGPEPTPVYGEKPSGKDPNTRIRQSDVGR 138

DB 562 RERYREEEKLQREDEKRR-----RQERQYRELEBLRQEQELDR 604

QY 139 DHRPQEQKKAGKLGKETILMQTLNTLSTPEKLAALCKKYALLLEE-HRNSQKMKLL 197

DB 605 KLREEQQLQER-----EELRQRQRERKLREEQQL--LRQEQELRQERKURREEQQL 659

QY 198 QKQSQQLVQEKDH-LRGEHSAVLARSKLSELCRELQHRNRSLSKBEQVQARREE---EK 253

DB 660 REEQQLQERKLRREEQ---LQERESEELRQERARKLRREEQQLRQEQELRQER 716

QY 254 RKEVTSHQVTLNDIQLOMEQHNENSKLRQENNEL-----AERLKKLIEQYELREEHDK 309

717 ERKLREEQQLRREEQQLRQ---ERDRKLREEQQLQSESEERLRRQERQEQQLRRER-DR 772

QY 310 VPKHKD-----LQQQLVDKQLQAQKMLKEAE-----EEHQEKDFLAKEA 350

DB 773 KFEDEEQQLQREERLRRQERARKLRREEQQLQREERLRRQERARKLRREEQQLQER 832

QY 351 VESORMC-----ELMKQEQTHLQKQALALYTEKPEEFONTLSKSESVFTTFKOE 398

DB 833 EERLRRQERARKLRREEQQLRQEQELRQERA---RKLREEQQLRQEQEQ---ELRQE 885

QY 399 MEKMTYKIKKL-----EKTTMYRWNSSNKALLEMAEKTVRDKLEGLQVKIQLEKL 454

DB 886 RDRKLREEQQLRQEQELRQERDRKLREEQQLQSESEERLRRQERER---KLREEQQL 942

455 CREALQTE-RNDLNKRAVD 471

943 LRREEQQLRREARKLRE 960

RESULT 4

RA50_ARCFU STANDARD; PRT; 886 AA.

AC Q29230.

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE DNA double-strand break repair rad50 ATPase.

GN RAD50 OR AF1032.

OS Archaeoglobus fulgidus.

OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;

OC Archaeoglobaceae; Archaeoglobus.

OX NCBI_TaxID=2234;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;

RX MEDLINE=98049343; PubMed=9389475;

RA Kleck H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D., Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T., Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

RA "The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus.";

RT Nature 390:364-370(1997).

RL Nature 390:364-370(1997).

CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The rad50/mrel complex possesses single-strand endonuclease activity and ATP-dependent double-strand-specific exonuclease activity. Rad50 provides an ATP-dependent control of mrel by unwinding and/or repositioning DNA ends into the mrel active site (By similarity).

CC -1- SUBUNIT: Forms a complex with mrel (By similarity).

CC -1- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AS001032; AAB90211.1; --

DR PIR; H69378; H69378.

DR TIGR; AF1032; --

DR HAMAP; MF 00449; -; 1.

DR InterPro; IPR003593; AAA_ATPase.

DR InterPro; IPR003439; ABC_transporter.

RESULT 6
DESP_HUMAN STANDARD; PRT; 2871 AA.
AC P15924; O75993; Q14189; Q9UNN4;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Desmoplakin (DP) (250/210 kDa paraneoplastic pemphigus antigen).
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM DPI).
RC TISSUE=Fore skin;
RX MEDLINE=92115697; PubMed=1731325;
RA Virata M.L.A., Wagner R.M., Parry D.A.D., Green K.J.;
RT "Molecular structure of the human desmoplakin I and II amino
terminus.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:544-548(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM DPI).
RC TISSUE=Fore skin;
RX MEDLINE=90153880; PubMed=1689290;
RA Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A., Wagner R.M.,
RA Angst B.D., Nilles L.A.;
RT "Structure of the human desmoplakins. Implications for function in
the desmosomal plaque.";
RL J. Biol. Chem. 265:2603-2612(1990).
RN [3]
RP SEQUENCE OF 1120-2871 FROM N.A. (ISOFORM DPI).
RC TISSUE=Fore skin;
RX MEDLINE=90153880; PubMed=1689290;
RA Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A., Wagner R.M.,
RA Angst B.D., Nilles L.A.;
RL J. Biol. Chem. 265:11406-11407(1990).
RN [4]
RP SEQUENCE OF 2854-2871 FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=20062965; PubMed=10594734;
RA Whitcock N.V., Ashton G.H., Dopping-Hepenstal P.J., Gratian M.J.,
RA Keane F.M., Eady R.A.J., McGrath J.A.;
RT "Striate palmoplantar keratoderma resulting from desmoplakin
haploinsufficiency.";
RL J. Invest. Dermatol. 113:940-946(1999).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=98012209; PubMed=9348293;
RA Kowalczyk A.P., Bornslaeger E.A., Borgwardt J.B., Palka H.L.,
RA Dhalwal A.S., Corcoran C.M., Denning M.F., Green K.J.;
RT "The amino-terminal domain of desmoplakin binds to plakoglobin and
clusters desmosomal cadherin-plakoglobin complexes.";
RL J. Cell Biol. 139:773-784(1997).
RN [6]
RP LIPIDATION.
RX MEDLINE=98316349; PubMed=9651377;
RA Marekov L.N., Steinert P.M.;
RT "Ceramide is bound to structural proteins of the human foreskin
epidermal cornified cell envelope.";
RL J. Biol. Chem. 273:17763-17770(1998).
RN [7]
RP VARIANT ARVD8 ARG-299.
RX MEDLINE=22285852; PubMed=12373648;
RA Rampazzo A., Nava A., Malacrida S., Beffagna G., Bauce B., Rossi V.,
RA Zambello R., Simionati B., Basso C., Thiene G., Towbin J.A.,
RA Danieli G.A.;
RT "Mutation in human desmoplakin domain binding to plakoglobin causes a
dominant form of arrhythmogenic right ventricular cardiomyopathy.";
RL Am. J. Hum. Genet. 71:1200-1206(2002).

[9]
RN RP VARIANTS SFVHS LYS-287 AND CYS-2366.
RX MEDLINE=21830938; PubMed=11841538;
RA Whitcock N.V., Wan H., Morley S.M., Garzon M.C., Kristal L., Hyde P.,
RA McLean W.H.I., Pulkkinen L., Uitto J., Christiano A.M., Eady R.A.J.,
RA McGrath J.A.;
RT "Compound heterozygosity for non-sense and mis-sense mutations in
desmoplakin underlies skin fragility/woolly hair syndrome.";
RL J. Invest. Dermatol. 118:232-238(2002).
CC - FUNCTION: Major high molecular weight protein of desmosomes.
CC Involved in the organization of the desmosomal cadherin-
plakoglobin complexes into discrete plasma membrane domains and in
the anchoring of intermediate filaments to the desmosomes.
CC - SUBUNIT: Homodimer.
CC - SUBCELLULAR LOCATION: Innermost portion of the desmosomal plaque.
CC - ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing, Named isoforms=2;
CC Name=DPI; Synonyms=Dp1;
CC IsoId=P15924-1, Sequence=Displayed;
CC Name=DPII; Synonyms=DP2;
CC IsoId=P15924-2, Sequence=VSP_005070;
CC - TISSUE SPECIFICITY: Isoform DPI is apparently an obligate
constituent of all desmosomes; Isoform DPII resides predominantly
in tissues and cells of stratified origin.
CC - DOMAIN: The N-terminal region is required for localization to the
desmosomal plaque and interacts with the N-terminal region of
plakophilin 1. The C-terminal region interacts with intermediate
filaments.
CC - PTM: Substrate of transglutaminase. Some glutamines and lysines
are cross-linked to other desmoplakin molecules, to other proteins
such as keratin, envoplakin, periplakin and involucrin, and to
lipids like omega-hydroxyceramide.
CC - DISEASE: Defects in DSP are the cause of familial arrhythmogenic
right ventricular dysplasia-8 (ARVD8) [MIM:607450], an autosomal
dominant disorder.
CC - DISEASE: Defects in DSP are a cause of striate palmoplantar
keratoderma II (PPK2, KPP2 or SPPK2), characterized by skin
thickening in the palms (linear pattern) and the soles (island-
like pattern) and flexor aspect of the fingers; and rarely by
abnormalities of the nails, the teeth and the hair.
CC - DISEASE: Defects in DSP are the cause of skin fragility-woolly
hair syndrome (SFVHS) [MIM:607655], an autosomal recessive
genodermatosis characterized by focal and diffuse palmoplantar
keratoderma, hyperkeratotic plaques on the trunk and limbs, and
woolly hair with varying degrees of alopecia.
CC - SIMILARITY: Contains 17 plectrin repeats.
CC - SIMILARITY: Belongs to the plakin or cytolinker family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M77830; AAA85135.1; -;
CC EMBL; AL031058; -; NOT ANNOTATED_CDS.
CC EMBL; J05211; AAA35766.1; -;
CC EMBL; AF139065; AAF19785.1; -;
CC PIR; A38194; A38194.
CC PDB; 1LM5; 31-JUL-02.
CC PDB; 1LM7; 31-JUL-02.
CC Genew; HGNC:3052; DSP.
CC MIM; 125647; -;
CC MIM; 607450; -;
CC MIM; 607655; -;
CC GO; GO:0005200; P:structural constituent of cytoskeleton; TAS.
CC GO; GO:0008444; P:epidermal differentiation; TAS.
CC InterPro; IPR001101; Plectrin_repeat.
CC InterPro; IPR002017; Spectrin.
CC Pfam; PF00681; Plectrin; 11.


```

FT DOMAIN 938 1507 23 x 23 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 938 961 2-1.
FT REPEAT 962 985 2-2.
FT REPEAT 986 1021 2-3.
FT REPEAT 1022 1044 2-4.
FT REPEAT 1045 1067 2-5.
FT REPEAT 1068 1090 2-6.
FT REPEAT 1091 1121 2-7.
FT REPEAT 1122 1144 2-8.
FT REPEAT 1145 1167 2-9.
FT REPEAT 1168 1197 2-10.
FT REPEAT 1198 1227 2-11.
FT REPEAT 1228 1250 2-12.
FT REPEAT 1251 1273 2-13.
FT REPEAT 1274 1296 2-14.
FT REPEAT 1297 1319 2-15.
FT REPEAT 1320 1342 2-16.
FT REPEAT 1343 1368 2-17.
FT REPEAT 1369 1391 2-18.
FT REPEAT 1392 1416 2-19.
FT REPEAT 1417 1439 2-20.
FT REPEAT 1440 1461 2-21.
FT REPEAT 1462 1484 2-22.
FT REPEAT 1485 1507 2-23.
FT VARSPLIC 1145 1197 Missing (in isoform Short).
FT VARSPLIC 1251 1273 /FTid-VSP 000847.
FT VARSPLIC 1399 1399 /FTid-VSP 000848.
FT CONFLICT 1549 1549 AA; 91173 MW; E72F99FF1326E54E CRC64;
SQ SEQUENCE 1549 AA; 91173 MW; E72F99FF1326E54E CRC64;

Query Match
Best Local Similarity 22.7%; Pred. No. 0.0032;
Matches 111; Conservative 108; Mismatches 194; Indels 77; Gaps 17;

Oy 6 QPEAGPGAGRPSQAAPAYAEAGPGSSQAPRPEGAQARTAGSGALRDVSELSQLED 65
Db 521 QREERKREGERERQYLEKVELQEEQLQREKREKREKQ-----YLEKVELQEE 574
Oy 66 ILSTYCVNNQGGPGEAGGAPBEPDAKSTYVARNGEPPTPVVYGEKPSKDPN 125
Db 575 QL-----QRQRQRQRQRERQYLEKVELQEEQLQREKREKREKRE 617
Oy 126 TE-----EIROSEVGDHRRRPOEKKAGLGEITLLMOTLNTLSTPEKLAALCKY 180
Db 618 RQYLEKVELQEEQVQREKREKREKREKREKREKREKREKREKREKREKREKRE 663
Oy 181 AELLERNSQKMLQKQSOVLQVQKDLH-RGEHSKAVLARSKLESLELQHRNRL 239
Db 664 EQLLEERERKREKREKREKREKREKREKREKREKREKREKREKREKREKREK 723
Oy 240 KEEGVORAREEERKREKREKREKREKREKREKREKREKREKREKREKREKRE 298
Db 724 QREKQLQREKREKREKREKREKREKREKREKREKREKREKREKREKREKRE 780
Oy 299 QYELREBHDVFKHDLQQLVDAKLAQQAQEMLEKAEERHQKDFLLKEAVESQRC 358
Db 781 QRE--EQLQREKREKREKREKREKREKREKREKREKREKREKREKREKRE 831
Oy 359 LMKQETHLQQLALYTESKEEP-----QNTLSKSEVFTTFKQEMKWTIKKLEK 412
Db 832 LEQLQEEELQRL-----DKRQFRDDQHQNEV-RNSRVYSKHNKES-----KOLD-D 881
Oy 413 TTYRSGESSKALLEMAEKRTVRDKLEGLGVK-----IQRLKCALQGTERRDLAK 467
Db 882 SWRESQFQDLPLQDEQEKREKREKREKREKREKREKREKREKREKREKREK 941
Oy 468 RVQDSAGGQ 477
Db 942 REEQLLKQ 951

```

RESULT 9

```

C190 DROME
ID C190 DROME STANDARD; PRT; 1690 AA.
AC Q9VJ5; O44929; Q8INY8; Q8MSD0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Restin homolog (Cytoplasmic linker protein 190) (Microtubule binding
protein 190) (d-CLIP-190).
GN CLIP-190 OR CG5020.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM A), FUNCTION, SUBCELLULAR LOCATION, AND
TISSUE SPECIFICITY.
RC STRAIN=Oregon-R; TISSUE=Embryo, and Ovary;
RX MEDLINE=98139549; PubMed=9472041;
RA Lantz V.A., Miller K.G.;
RT "A class VI unconventional myosin is associated with a homologue of a
microtubule-binding protein, cytoplasmic linker protein-170, in
neurons and at the posterior pole of Drosophila embryos.";
RL J. Cell Biol. 140:897-910(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Bencos P.V., Berman B.P., Brokstein P., Brottier P.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foeller C., Gabrielian A.R., Garg N.S., Galbart W.M., Glasser K.,
Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Stadling A.C., Stapleton M., Strong R., Sun E.,
Swirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasarman D.A., Weinstock G.M., Weissenbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Zhu X., Smith H.O.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhao Q., Zheng L.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [3]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,

```


CC -I- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be
CC split further into 2 globular subfragments (S1) and 1 rod-shaped
CC subfragment (S2).
CC -I- SIMILARITY: Contains 1 IQ domain.
CC -I- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL: V00430; CAA23712.1; --
CC EMBL: J02714; AAA48972.1; --
CC PIR: A29320; A29320.
CC HSP: P13538; 2MYS.
CC InterPro: IPR000048; IQ region.
CC InterPro: IPR001609; myosin head.
CC InterPro: IPR004009; myosin_N.
CC InterPro: IPR002928; myosin_tail.
CC Pfam: PF00612; IQ; 2.
CC Pfam: PF00663; myosin_head; 1.
CC Pfam: PF02736; Myosin N; 1.
CC Pfam: PF01576; Myosin Tail; 1.
CC PRINTS: PR00193; MYOSINHEAVY.
CC ProDom: PD000355; myosin_head; 1.
CC SMART: SM00015; IQ; 1.
CC SMART: SM00242; MYSC; 1.
CC PROSITE: PS00096; IQ; 1.
CC Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC Calmodulin-binding; ATP-binding; Methylation; Multigene family.
CC DOMAIN 1 785
CC MYOSIN HEAD-LIKE.
CC DOMAIN 786 815
CC IQ.
CC COILED COIL (POTENTIAL).
CC NP_BIND 844 1940
CC ATP.
CC DOMAIN 179 186
CC ACTIN-BINDING.
CC DOMAIN 660 682
CC METYLATION (TRI-) (POTENTIAL).
CC MOD_RES 130 130
CC VARIANT 379 379
CC T -> A (IN REF. 2).
CC CONFLICT 1547 1547
CC ERA -> GRT (IN REF. 2).
CC CONFLICT 1913 1915
CC ERA -> GRT (IN REF. 2).
CC SEQUENCE 1940 AA; 222816 MW; C34833D75B04DF2 CRC64;
Query Match 9.1%; Score 245; DB 1; Length 1940;
Best Local Similarity 21.8%; Pred. No. 0.0072;
Matches 114; Conservative 87; Mismatches 180; Indels 142; Gaps 16;
33 SCAPKPCGAQARTAQSGALRDVSEELSRQLEDILTVCVDNNGGPGEDGAGGEPAPPE 92
1093 SQIOSKIDEDQALGNQ---LQKKIKELQARIEEL-----EEIEAERTSRA 1135
93 DAKRSRTVYVANGPEPTPVVYVGKESKGPENTEEIRQSDVEGDRHRRPQEKKAGKL 152
1136 KAEHRADLSRELE-----ISERLEAGGNTAAQI-----DMNKKREAFQKVRD 1182
153 GKEITLLMOTLNTSTPEKLAALCKY-----ALLEHRNSQKMLQKQSQLYQEK 208
1183 LEEATL-----QHEATAALRKGHADSTADVGEQIDLRVQKLEKEKSELQMEI 1233
209 DILRGESKAVLARSKLSLRCLEL-----QRNRSLEKEGVQVQAR----- 248
1234 DLLASNMESVSKANLEKYCRSLDQSLSEIKTKREEQORTINDISAKRLQTESGEYS 1293
249 -----EEBKKEVTSFH-FQVTLNDIQLQMEQHN 276
1294 RQVEKDKALISOLSRGKQAFQIIBELKRLHLEETKAKKCPAHALQSAHDCDLLREQYE 1353
277 ER-----NSKLR-----QENNELAEELKLEIYVELREEHIDKVP 311
1354 BEQKAGELQRLSVANSEVAGWTKYETDALQRTSELEBAKKLQALQDAEEHV----- 1409

QY 312 KHKDLOQQQLVDAKLQQAQOMLKEAEERHQEKDFLLKEAVESQRCMELMKQOETHLKQOL 371
Db 1410 -----EAVNSKCA-----LEKTKQRLQNEVEDLMIDVERSNACAAALDKQKQNFDKIL 1458
QY 372 ALYTEKFEFFONTLSKSSVFTFFQOEMKMTKKIKLSEKTTMYRSRWESSNKALLEWA 431
Db 1459 SEWKKYEEVTEAELEASQKESLSLSTELPFM-----KNAYEESLDHLLETUKRENKILQOEI 1514
QY 432 ESKTVRDKLEGLQVQKIQRLKLCRALQTERNDLNKRVQDLQA 474
Db 1515 SLDL---EQIAGGKA-IHELEKVKQIQEKEKSELQTALEBAEA 1554
RESULT 14
RBLI HUMAN STANDARD: PRT; 1410 AA.
AC Q9P2E9; O75300; O968S2; Q9BWPL; Q9H476;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ribosome-binding protein 1 (Ribosome receptor protein) (180 kDa
DE ribosome receptor homolog) (ES/130 related protein).
GN RRBPI OR KIAA1398.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=98290552; PubMed=9628588;
RA Langley R. Leung E., Morris C., Berg R., McDonald M., Weaver A.,
RA Parry D., Ni J., Su J., Gentsz R., Spour N., Krissansen G.W.;
RT "Identification of multiple forms of 180-kDa ribosome receptor in
RT human cells.";
RL DNA Cell Biol. 17:449-460(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RX TISSUE=Brain;
RC MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirotsawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavridis G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.V., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhali P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lavelle S.,
RA Levaeslaih M.H., Leverhwa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.B., McConachie L.J., McWay K., McKerray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Thorpe A.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;

QY 515 EASGQTGPQBPSTA 528
Db 1109 DLASKLREABETQS 1122

RESULT 15
SCPI1 MOUSE
ID SCPI1_MOUSE STANDARD; PRT; 993 AA.
AC Q62209; 009205; P70192; Q62329;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, last sequence update)
DT 15-DEC-1998 (Rel. 37, last annotation update)
DE Synaptonemal complex protein 1 (SCP-1 protein).
GN SCYP1 OR SCPI1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RN SEQUENCE FROM N.A.
RP STRAIN=CBA; TISSUE=Testis;
RX MEDLINE=96004899; PubMed=7548215;
RA Sage J., Martin L., Cuzin F., Rassoulzadegan M.;
RT "cDNA sequence of the murine synaptonemal complex protein 1 (SCP1).";
RL Biochim. Biophys. Acta 1263:256-260(1995).
RN (2)
RN SEQUENCE FROM N.A.
RP STRAIN=Swiss; TISSUE=Testis;
RX Kerr S.M., Taggart M.H., Lee M., Cooke H.J.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBAJ databases.
RN (3)
RN SEQUENCE OF 1-149 FROM N.A.
RP STRAIN=C57BL/6;
RA Sage J., Li Y., Martin L., Mattei M.-G., Guenet J.-L., Liu J.G.,
RA Hoog C., Cuzin F., Rassoulzadegan M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBAJ databases.
RN (4)
RN SEQUENCE OF 95-787 FROM N.A.
RP STRAIN=ICR; TISSUE=Testis;
RA Tsuchida J., Nishina Y., Nozaki K., Nishimune Y.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBAJ databases.
CC -!- FUNCTION: Major component of the transverse filaments of
CC synaptonemal complexes (SCS), formed between homologous
CC chromosomes during meiotic prophase.
CC -!- SUBCELLULAR LOCATION: Nuclear. In tripartite segments of
CC synaptonemal complexes, between lateral elements in the nucleus.
CC Found only where the chromosome cores are synapsed. Its N-terminus
CC is found towards the centre of the synaptonemal complex while the
CC C-terminus extends well into the lateral domain of the
CC synaptonemal complex (By similarity).
CC -!- DOMAIN: Consists of an alpha-helical stretch of 700 AA residues,
CC flanked by N- and C-terminal globular domains. The C-terminal
CC domain has DNA-binding capacity (by similarity).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; Z38118; CAAS6262.1; -
CC EMBL; L41069; AAA64514.1; ALT_INIT.
CC EMBL; U62864; AAC53335.1; -
CC EMBL; U62860; AAC53335.1; JOINED.
CC EMBL; U62861; AAC53335.1; JOINED.
CC EMBL; U62862; AAC53335.1; JOINED.
CC EMBL; U62863; AAC53335.1; JOINED.
CC EMBL; D88539; BAA13639.1; -
CC PIR; S49461; S49461.
CC MGD; MGI:105931; Sycp1.

GO; GO:0000795; C:synaptonemal complex; IDA.
DR InterPro; IPR008827; SCP-1.
DR Pfam; PF05483; SCP-1; 1.
KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT DOMAIN 12 97 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 104 815 COILED COIL (POTENTIAL).
FT DOMAIN 114 117 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 697 700 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 898 901 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 978 986 ARG/LYS-RICH (BASIC).
FT CONFLICT 527 527 F -> L (IN REF. 2).
SQ SEQUENCE 993 AA; 115962 MW; 1A4FA790D64FAFE6 CRC64;

Query Match 9.0%; Score 243.5; DB 1; Length 993;
Best Local Similarity 22.9%; Pred. No. 0.0043;
Matches 124; Conservative 91; Mismatches 183; Indels 143; Gaps 23;

QY 34 QAPRKEPGAQAPTAQSGALRDVSELSROLED-----ILSYCVND 73
Db 132 QENRKLIIEAQRKAIQI--LQFENEKYSKLKEEIQENKDLKENNATIEHCNLLKFC-- 187
QY 74 NNQGFGEDEGAQGEPAEPEDAESKRTYVARNGEPPTPVVYGEKPSKGDPTNEEQRSD 133
Db 188 -----ARSAEKTNKYEYERETROVYVDINSNIEMILAF-----EELRVQA 229
QY 134 EVG-----DRDHERPQ--EKKKAGLGKGIITLLMOTLTLSTPBEKLAALCKKYAEL 183
Db 230 ENARLEHGFKLKEDHEKIQHLEBEYQKRVNNKQVSELLIQSAEKENM-----KLTFL 285
QY 184 LEEHRNSQKQW----KLLQKQSQVQEKDHLRGEHSKAVLARSKLESLECRELQRH---N 236
Db 286 LEESRDKANQLEKTKLQDENLKLSEKNDHL-----TSELEDIKMSQMSMTQ 335
QY 237 RSLKEE-----GYORAREBEERKEKVTSHFQV-----TLNDIQLOMEQHN 276
Db 336 KALEEDIQIATKTSIQLTSEYKRAQMEELNCAKTHSFVVTTELKATCTTSELTEQRL 395
QY 277 ERNS-----KLROENMELAEELKGLIQQYELREHIDKVFKHQLOQLOQVDAKLQQA 328
Db 396 EKEDQLKLITVELQKSNEL--EEMTKFNKKEVELEELKNILAE---DQKLDEK-KQV 450
QY 329 QEMLKAEERHQRKDFLL-----KEAVSQRMCELKQKQETH-LKQQLALYTE-KKEEFQ 382
Db 451 EKLAELQKKEQ-ELTFLLETREREKVDLQEQVTVTKTSBQHYLKQVEEMKTELEKELK 509
QY 383 NT-----LSKSSEVFTTFQKMEKMTKTKLEKETTMY 416
Db 510 NTELTASCDMLLENKFFQVQESDMALELKKHQEDIIINCKQERLLKQIENLEKEMHL 569
QY 417 RSRWESNKKALLEMAEP-KTVDRKELE---GLQVKQRLEKCRALQTERNDLNKRVQDL 472
Db 570 RDELESVRKEPFIOQGDVEVKCKLDKSEENARSIECEVLKKEQMKILESKNNLKKQVENK 629
QY 473 S 473
Db 630 S 630

RESULT 16
PLE1 CRIGR
ID PLE1 CRIGR STANDARD; PRT; 4473 AA.
AC Q9J155;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 15-MAR-2004 (Rel. 43, last annotation update)
DE Plectin 1 (PLTN) (PCN) (300-kDa intermediate filament-associated
DE protein) (IPAP300) (Fragment).
GN PLE1
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.


```
QY 488 RPEGQAQ-PSSPRVTEAPCVGPSTPESG 518
Db 2543 ATQAAAKALPNGRDALDGSMEVEPEHAFEG 2574

RESULT 17
ID RBBI_CANFA STANDARD; PRT; 1534 AA.
AC Q28298;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosome-binding protein 1 (180 kDa ribosome receptor) (RRP).
GN RBP1 OR P180.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RC TISSUE=Kidney;
RX MEDLINE=95310363; PubMed=7790375;
RA Wanker E.E., Sun Y., Savitz A.J., Meyer D.I.;
RT "Functional characterization of the 180 kDa ribosome receptor in
RL J. Cell Biol. 130:29-39(1995).
CC -1- FUNCTION: Acts as a ribosome receptor and mediates interaction
CC between the ribosome and the endoplasmic reticulum membrane.
CC -1- SUBCELLULAR LOCATION: Type III membrane protein. Endoplasmic
CC reticulum.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; X87224; CAA60676.1;
CC PIR; A56734; A56734.
CC DR GO; GO:0030176; C:integral to endoplasmic reticulum membrane; IDA.
CC DR GO; GO:0030376; P:ribosome receptor activity; IDA.
CC DR GO; GO:0007165; P:signal transduction; IDA.
CC DR InterPro; IPR007794; Rib rec.KP reg.
CC DR Pfam; PF05104; Rib rec.KP reg. 1.
CC KW Transport; Protein transport; Translocation; Endoplasmic reticulum;
CC Transmembrane; Repeat; Alternative splicing.
CC DOMAIN 1 7 LUMENAL (POTENTIAL).
CC FT TRANSMEM 8 28 POTENTIAL.
CC FT DOMAIN 29 1534 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 197 736 54 X 10 AA APPROXIMATE TANDEM REPEATS OF
CC FT N-Q-G-K-K-A-E-G-A-P.
CC SQ SEQUENCE 1534 AA; 164586 MW; B343BCF12656F3C5 CRC64;

Query Match 9.0%; Score 242; DB 1; Length 1534;
Best Local Similarity 20.4%; Pred. No. 0.0075;
Matches 153; Conservative 114; Mismatches 222; Indels 260; Gaps 30;

QY 1 KSSPGQPEAGPE -----GAQRPQAPAVEAEGPGSSOAP 36
Db 710 KXSEGSFNGKKVDASANSQKASAPGKAGVQSQEAPKQAPAKKSGSKGKEP 769
QY 37 RKP -----EQAQAR -----TAQSGALRDVSEB ----- 58
Db 770 GPDSDSPLYPKTLVSTVGMVFNBSGAQLBILSEKAGVTDYTHKATQKGDPAI 829
QY 59 LSRQLSD-----ILSTYCDVNNQGGPGDGAQSPAPDEAKSRT---YVAENGPEPEFV 112
Db 830 LKRLQEEKEKLAT-----SQEDAAYAKSKLRVYNKELAEAKA 869
QY 113 VTGEKPSK-----GPDNTEISQSDVGDHRRRQEKKAK-----G 151
```

RA Suda M., Misumi Y., Fujiwara T., Nishioka M., Ikehara Y.;
 RT "Molecular cloning and sequence analysis of a human 372-kDa protein
 RL localized in the Golgi complex";
 CC Biochem. Biophys. Res. Commun. 205:1399-1408(1994).
 CC -!- FUNCTION: May participate in forming intercisternal cross-bridges
 CC of the Golgi complex.
 CC -!- SUBUNIT: Homodimer; disulfide-linked.
 CC -!- SUBCELLULAR LOCATION: Golgi; membrane-associated.
 CC -!- DISEASE: Antigen in chronic rheumatoid arthritis and in the
 CC autoimmune disease Sjogren's syndrome.
 CC -!- SIMILARITY: Belongs to the golgin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to licenses@isb-sib.ch).
 CC

EMBL; X75304; CAA53052.1; -;
 DR EMBL; D25542; BAA05025.1; -;
 DR PIR; A56539; A56539.
 DR PIR; I52300; I52300.
 DR Genew; HGNC:4429; GOLGB1.
 DR MIT; 602500; -;
 DR GO; GO:0000139; C:Golgi membrane; TAS.
 DR GO; GO:0005795; C:Golgi stack; TAS.
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR GO; GO:0007030; P:Golgi organization and biogenesis; TAS.
 KW Golgi stack; Antigen; Coiled coil; Transmembrane.
 FT DOMAIN 1 3235 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 3236 3256 POTENTIAL.
 FT DOMAIN 3257 3259 LUMENAL (POTENTIAL).
 FT DOMAIN 48 593 COILED COIL (POTENTIAL).
 FT DOMAIN 677 1028 COILED COIL (POTENTIAL).
 FT DOMAIN 1062 1245 COILED COIL (POTENTIAL).
 FT DOMAIN 1301 1779 COILED COIL (POTENTIAL).
 FT DOMAIN 1828 3185 COILED COIL (POTENTIAL).
 FT DOMAIN 2420 2423 POLY-GLU.
 FT DOMAIN 2993 2996 POLY-SER.
 FT CONFLICT 1 39 MISSING (IN REF. 3).
 FT CONFLICT 215 215 A -> AQLSSM (IN REF. 3).
 FT CONFLICT 1765 1765 D -> G (IN REF. 3).
 FT CONFLICT 2950 2950 H -> D (IN REF. 3).
 SQ SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178DD CRC64;

Query Match 9.0%; Score 242; DB i; Length 3259;

Best Local Similarity 20.4%; Pred. No. 0.016;
 Matches 121; Conservative 99; Mismatches 168; Indels 206; Gaps 22;

QY 43 QARTAGSGLRDVSELSRQLEDILSTYC-----VDNNQGGPGEDGAGGPAPBDEAK 96
 DB 1743 QSLMSEKDSLSSEVQDLKHQIEDNYSKQANLEATEKHNDQNTVTEEGTQSPGTEHDS 1802
 QY 97 -STYVARGPEPT-----PVYGEKPSKGDPTNTEIRQSDVGDHRRPPOEKKAK 150
 DB 1803 LSWSTRPTCSSEVPSAKSANPAV--SKDFSSHDEINNYLIQDQLKRIAGLEEKQK 1860
 QY 151 GLAKEI-----TLTMO-----TLNTLS-----TPBE--- 171
 DB 1861 EFSQTLNENKNTLLSQISTKQGLKMLQBEVTKMLNQLQIQEELSRVTKLXTAREKD 1920
 QY 172 -----KLAL-----CKYA-----ELLE-EHRNSQKMKLLQKQSOVLQVKD 209
 DB 1921 DLBERLNLQALNGLSNGYICQDVDTAQIKNLESEMKNLKKCVSELEEEKQLVKEKT 1980
 QY 210 HLAGHESKAVLAR-----SKLSLCELOHRNLSK----- 240
 DB 1981 KYSEIRKYLEKIQAQKEPGNKSNAKELQELLEKEQEVQLQKQICIRYQEKISALER 2040
 QY 241 -----EGVQARAREERKRVKVTSHFQVTLNDIQLQMEQNRNSK 281

DB 2041 TVKALEFVOTESQOLEITKENLAQAVEHKKQAELASFKVLLDDTQSEARVLADNLK 2100
 QY 282 LROENNELASRLKGLLEOVELREHHDVFKFKKOLQOOLVDKQLQQAQEMLKEAEEHQR 341
 DB 2101 LKGLQSNKSEVSKQKQ-----KQEDLERR-----LQAEKXK-- 2134
 QY 342 EKDFLLKAVESQRMCELMKQKQTHLQKQALYTFKFEFQNTLSKSSVFTTFKQEMEK 401
 DB 2135 -----LKEKNMQEKLDALEERKVLHEETIG-----BIQVTLNKKCKEVLQQLQENLDS 2182
 QY 402 -----MTKKIKKLEKTTMY-----RSRWESSNKALLEMAEKTVRDKS----- 440
 DB 2183 TVTQLAAPYKSMSSLODDRDRVIDEAKKWERKESDAIQSKEE-ILKEDNCSVLKQDLR 2241
 QY 441 -----LEGQVQKIFQRL-----KLCRALQTERNDLMKRVQD 471
 DB 2242 QMSIRMEELKINISRLHDKQIWESRAQTEVQLQKQKVCDTLQENKELLSQLEE 2295

RESULT 19

CENE_HUMAN

ID CENE_HUMAN STANDARD; PRT; 2663 AA.

AC Q02224;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Centromeric protein 2 (CENP-E protein).

GN CENPE

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RA MEDLINE=93024922; PubMed=1406971;

RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;

RT "CENP-E is a putative kinetochore motor that accumulates just before

RT mitosis.";

RL Nature 359:536-539(1992).

RN (2)

RP CHARACTERIZATION.

RA MEDLINE=95196755; PubMed=7889940;

RA Thresher D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;

RT "Mitotic Hela cells contain a CENP-E-associated minus end-directed

RT microtubule motor.";

RL EMBO J. 14:918-926(1995).

RN (3)

RP CHARACTERIZATION.

RA MEDLINE=98437347; PubMed=9763420;

RA Chan G.K.T., Schaar B.T., Yen T.J.;

RT "Characterization of the kinetochore binding domain of CENP-E reveals

RT interactions with the kinetochore proteins CENP-F and hBUBR1.";

RL J. Cell Biol. 143:49-63(1998).

RN (4)

RP FARNESYLATION.

RA MEDLINE=20459117; PubMed=10852915;

RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.;

RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E

RT and CENP-F and alter the association of CENP-E with the

RT microtubules.";

RL J. Biol. Chem. 275:30451-30457(2000).

CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE

CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE

CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT

CC AND/OR SPINDLE ELONGATION.

CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.

CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING

CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS

CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.

CC -!- SIMILARITY: Belongs to the kinesin-like protein family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z15005; CAA78727.1; --
DR PIR: S28261; S28261.
DR HSSP: P17119; 3XAR.
DR Genew: HGNC:1856; CENPE.
DR GK: Q02224; --
DR MIM: 117143; --
DR GO: GO:0005699; C:kinetochore; TAS.
DR GO: GO:0005334; C:nucleus; TAS.
DR GO: GO:0008350; F:kinetochore motor activity; TAS.
DR GO: GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR GO: GO:0007079; P:mitotic chromosome movement; TAS.
DR GO: GO:0007080; P:mitotic metaphase plate congression; TAS.
DR InterPro: IPR001752; kinesin_motor.
DR Pfam: PF00225; kinesin; 1.
DR PRINTS: PR00380; KINESINHEAVY.
DR SMART: SM00129; KISC; 1.
DR PROSITE: PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE: PS00067; KINESIN MOTOR DOMAIN2; 1.
DR Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere; Lipoprotein; Prenylation.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
FT LIPID 2660 2660 S-farnesyl cysteine.
SQ SEQUENCE 2663 AA; 312087 MW; CFCF1380C8C8CB8 CRC64;

Query Match 8.9%; Score 240; DB 1; Length 2663;
Best Local Similarity 20.2%; Pred. No. 0.015;
Matches 130; Conservative 101; Mismatches 176; Indels 238; Gaps 25;

QY 55 VSELSHQLEDILSTYCVNNQGGPGEAGG-----EPASPDAKSRVYA----- 102
DB 987 ISESVRNILK-----MEENTGETDFQKGVGIDKKODLEAKNTQTLTADVKNEII 1039
QY 103 -----RNGSEPPTVYVGEKPSGSD-----PNTETIR-QSDRVGDRDH 140
DB 1040 EQQRKIFSLIOEKNELOQMLESVTAKEQKLTDLKENIEMTIENQEEERLLGLDKKQOE 1099
QY 141 RRPQKKKA-KGLGKEITLLMTLNTLSTPEKL----- 173
DB 1100 IVAQEKNAIKKEG-----LSRTCDRLAEVSEKLEKSKSQQLQKQQLLVQVEMSEMOK 1155
QY 174 -----AALCKKYAELEEHNSOKMKLQKQSQQLVQOE 207
DB 1156 KINEIENLKNELKNELTLEHMETERLELAQKLANEYEVKSIKRRKVKELQKSPETS 1215
QY 208 XDLRLG-----EHSKAV--LARSKLESLC-----RELQRHN 236
DB 1216 RDHLGTYREIATGLQTKELKLAHILKHEQTHIDELRESVEKTAQIINTQDLKSH 1275
QY 237 RSLKEGVQVQARBEKKKQV--TSHFQVTLNDIQLQHEQHNENS-----KLRENMLA 290
DB 1276 TGLQRE-IPVLHEEQELLPVKKVSETQETVNELELLTEQSTTDDSTTLARIEMERLRN 1334
QY 291 FELKXLIQYELREHIDKV-----FKKDLQQLQVLD--AKLQQAQ----- 329
DB 1335 EKFSQSEIKSLTKERONLKTIKALRVKDDQLKEHIRETLAKIQSSQKQESLNWKE 1394
QY 330 -----EVL-----KEAEERHQREKDFLKEAVESQRMCEL 359
DB 1395 KDNETTKIVSEMQPKDSALLRIEIEMLGLSKQLQSHDEMKS-VAKEKXDDLQRLQEV 1453
QY 360 MKQETHLKOOL-----ALYTKFRFQ-----NTLSKSEVFTTFKQ 397
DB 1454 LQESDQLKENIKETIAKHLETEBEKRVKVAHCLKEQETINELRVNLSEKETEISTQKQ 1513

QY 398 ---EMEKTKIKKL-EKETTM---YRSRWESSKALLEMAEKTVDKLEGLQVYKI-- 448
DB 1514 LEAINDKQNKQIQIYEEKEQLNIQISVQENVNKLQPKHEKRAKDSALQSISSOMLE 1573
QY 449 -----ORLEKLCRAIQTERNDINRQVQDLSA 474
DB 1574 LTNRLQESQEBTQIMIKKEENKRVQEAQIERDQKENTKEIVA 1618

RESULT 20
TRHY_HUMAN STANDARD; PRT; 1898 AA.
ID TRHY_HUMAN
AC Q07283;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Trichohyalin.
GN THH OR TRHY OR THH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:93380194; PubMed:7685034;
RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
RA Steinert P.M.;
RT "The structure of human trichohyalin. Potential multiple roles as a
RT functional EF-hand-like calcium-binding protein, a cornified cell
RT envelope precursor, and an intermediate filament-associated (cross-
RT linking) protein";
RT J. Biol. Chem. 268:12164-12176(1993).
RN [2]
RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX MEDLINE:93315897; PubMed:7686953;
RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;
RA "Trichohyalin: a structural protein of hair, tongue, nail, and
RA epidermis";
RL J. Invest. Dermatol. 101:658-718(1993).
CC -I- FUNCTION: Intermediate filament-associated protein that associates
CC in regular arrays with keratin intermediate filaments (KIF) of the
CC inner root sheath cells of the hair follicle and the granular
CC layer of the epidermis. It later becomes cross-linked to KIF by
CC isodipeptide bonds. It may serve as scaffold protein, together
CC with involucrin, in the organization of the cell envelope or even
CC anchor the cell envelope to the KIF network. It may be involved in
CC its own calcium-dependent postsynthetic processing during terminal
CC differentiation.
CC -I- SUBUNIT: Monomer (probable).
CC -I- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as
CC the inner root sheath (IRS) of hair follicles and medulla, and in
CC the filiform papillae of dorsal tongue epithelium (Probable).
CC -I- DEVELOPMENTAL STAGE: Expressed during late differentiation of
CC the epidermis.
CC -I- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand
CC calcium-binding domains. Domains 2-4, 6, and 8 are almost
CC entirely alpha-helical, configured as a series of peptide repeats
CC of varying regularity, and are thought to form a single-stranded
CC alpha-helical rod stabilized by ionic interactions. Domain 6 is
CC the most regular and may bind KIF directly by ionic interactions.
CC domains 5 and 7 are less well organized and may induce folds in
CC the molecule. Domain 9 contains the C-terminus, conserved among
CC different species.
CC -I- PTM: Substrate of transglutaminase. Some 200 arginines are
CC probably converted to citrullines by peptidylarginine deiminase.
CC -I- SIMILARITY: In the N-terminal section; belongs to the S-100
CC family.
CC -I- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=F1B; TISSUE=Liver;
 RX MEDLINE=95115033; PubMed=7815459;
 RA Wang R., Sole M.J., Cukerman E., Liew C.-C.;
 RT "Characterization and nucleotide sequence of the cardiac alpha-myosin
 heavy chain gene from Syrian hamster."
 RL J. Mol. Cell. Cardiol. 26:1155-1165(1994).
 RN [2]
 RP SEQUENCE OP 962-1935 FROM N.A.
 RX MEDLINE=88247788; PubMed=3380703;
 RA Jandreski M.A., Sole M.J., Liew C.-C.;
 RT "Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin
 heavy chain."
 RL Nucleic Acids Res. 16:4737-4737(1988).
 CC -|- FUNCTION: Muscle contraction.
 CC -|- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
 heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
 and 2 regulatory light chain subunits (MHC-2).
 CC -|- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -|- DOMAIN: The rodlike tail sequence is highly repetitive, showing
 cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
 characteristic for alpha-helical coiled coils.
 CC -|- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
 meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be
 split further into 2 globular subfragments (S1) and 1 rod-shaped
 subfragment (S2).
 CC -|- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -|- SIMILARITY: Contains 1 IQ domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 DR EMBL; L12104; AAA62313.1; --
 DR EMBL; X07273; CAA30256.1; --
 DR F1R; I48153; 148153.
 DR HSP; P08799; IMND.
 DR InterPro: IPR000048; IQ region.
 DR InterPro: IPR001609; myosin head.
 DR InterPro: IPR004009; Myosin N.
 DR InterPro: IPR002928; Myosin_tail.
 DR Pfam; PF00612; IQ; 2.
 DR Pfam; PF00663; myosin head; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS50096; IQ; 1.
 DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylatoin; Multigene family; Calmodulin-binding.
 FT DOMAIN 1 779
 FT MYOSIN HEAD-LIKE.
 FT DOMAIN 780 809
 FT IQ.
 FT DOMAIN 839 1934
 FT COILED COIL (POTENTIAL).
 FT NP_BIND 177 184
 FT ATP.
 FT ACTIN-BINDING.
 FT DOMAIN 654 676
 FT MYOSIN-BINDING.
 FT MOD_RES 128 128
 FT METHYLATION (TRI-) (POTENTIAL).
 FT CONFLICT 966 966
 FT D -> E (IN REF. 2).
 FT CONFLICT 978 978
 FT T -> TE (IN REF. 2).
 FT CONFLICT 986 986
 FT E -> Q (IN REF. 2).
 FT CONFLICT 1008 1014
 FT DLCAED -> ALCAET (IN REF. 2).
 FT CONFLICT 1057 1057
 FT D -> Y (IN REF. 2).

FT CONFLICT 1060 1060 L -> V (IN REF. 2).
 FT CONFLICT 1095 1095 D -> N (IN REF. 2).
 FT CONFLICT 1217 1217 E -> D (IN REF. 2).
 FT CONFLICT 1271 1271 D -> N (IN REF. 2).
 FT CONFLICT 1327 1327 T -> A (IN REF. 2).
 FT CONFLICT 1327 1327 C -> R (IN REF. 2).
 FT CONFLICT 1358 1358 L -> V (IN REF. 2).
 FT CONFLICT 1504 1504 L -> V (IN REF. 2).
 FT CONFLICT 1537 1537 M -> L (IN REF. 2).
 FT CONFLICT 1556 1556 N -> K (IN REF. 2).
 SQ SEQUENCE 1934 AA; 222928 MW; FDBAC58310B0B57D CRC64;
 Query Match 8.8%; Score 237; DB 1; Length 1934;
 Best Local Similarity 23.3%; Pred. No. 0.015;
 Matches 129; Conservative 88; Mismatches 208; Indels 128; Gaps 20;
 QY 12 EGAQRRPSQAAPAVEAEGPGSSQAPRKPQCAQARTAQSGALRDVSELSQLDILSTYC 71
 DB 1034 EGSLEQEKVMDLE-----RAKKLEGDLKLTQBS--IMDLEND-KQOLEKLLKDD 1083
 QY 72 VDNNGGPGEDGAGCPAPEDA-----EKSTTYVAR-----NGEPETPVVYGEKEP 119
 DB 1084 FELN-----ALNARIEDEQALGSLQKGLKELQARIIEELBELEAERTAKVEKLR 1135
 QY 120 SKGDPNTEIROS-DEVG-----DRHRRPOKKKAGLGEITILLMOTLNTLSTPEE 171
 DB 1136 SLSRELEIEISERLEHAGGATSVQIENMKREAEFFQWRDLLEATL-----QHEA 1186
 QY 172 KLAALCKKY-----ABLEEHRNSQOMKLLQKOSQLVQSKDHLRGHSHAVLARSKLES 227
 DB 1187 TAAALRKKHADSVAELEGQIDNLRQVKLEKSEKSEFKLEDDVTSMNQIIRAKANLEK 1246
 QY 228 LCR-----ELQRHN-----RSLKE----- 241
 DB 1247 MCTLEQDMEHRSKAEETORSVNDLTSQRAKQOTENGELSRLDKEALISQITRGKLT 1306
 QY 242 -----EGVQARAREEEXRKEVTSH-FQVTLNDIQLQMEQHNERNK-----LRQENME 288
 DB 1307 YTQOLEDLKQLESEVAKNTLAHALQASARHDCDLLREQYEEETEAKAELQCVLKANSK 1366
 QY 289 LAELKKLIEQYELREHIDKVPKDLQOOLVDKALQQAQE-----MLKBAEERHOR 341
 DB 1367 VAQWRTYETDAIQRTBELSEA--KKKLAORLQDA--BEAVEAVNAKSSLEKTKHRLQN 1422
 QY 342 EKDFLLKAVESQRCBMLMQOETHLKQALAYTEKFEFPQNTLSKSEVFTTTPQENMEK 401
 DB 1423 EIEDLWVDVRSNAAAALDKQRNPKILAEWKQYEESSQSESSQKARSLSLTEL-- 1480
 QY 402 MTKKIKKLEKETTYRGRWSSNKALLEMAEKTVRDKLEGLQVQIKRLEKLCALQTE 461
 DB 1481 --FKLKNAYEESLEHLETFKRENKQLQFEISDLT---EQLGSTGKSHHELEKIRKQLEAE 1535
 QY 462 RNDLNKRVQDLQA 474
 DB 1536 KNELOSALEBAEA 1548
 RESULT 25
 MYHD HUMAN
 ID MYHD HUMAN STANDARD; PRT; 1938 AA.
 AC Q9UKX3; O95252;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin heavy chain, skeletal muscle, extraocular (MyHC-eo).
 GN MYH13.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Extraocular muscle;
 RC MEDLINE=9911869; PubMed=10388558;
 RX

FT	657	679	ACTIN-BINDING.
FT	759	773	ACTIN-BINDING.
FT	129	129	METHYLATION (TRI-)
FT	MOD RES	88	Q -> E (IN REF. 1).
FT	CONFLICT	88	Q -> P (IN REF. 1).
FT	574	574	Q -> G (IN REF. 1).
FT	CONFLICT	608	T -> A (IN REF. 1).
FT	CONFLICT	744	M -> I (IN REF. 1).
FT	CONFLICT	790	V -> A (IN REF. 1).
FT	CONFLICT	1014	S -> T (IN REF. 1).
FT	CONFLICT	1021	A -> V (IN REF. 1).
FT	CONFLICT	1101	A -> S (IN REF. 1).
FT	CONFLICT	1290	W -> C (IN REF. 1).
FT	CONFLICT	1373	K -> N (IN REF. 1).
FT	CONFLICT	1533	L -> M (IN REF. 1).
FT	CONFLICT	1540	KL -> NV (IN REF. 5).
FT	CONFLICT	1577	EQ -> DR (IN REF. 1).
FT	CONFLICT	1705	E -> D (IN REF. 1).
FT	CONFLICT	1733	A -> S (IN REF. 2).
FT	CONFLICT	1734	T -> S (IN REF. 1).
FT	CONFLICT	1737	T -> H (IN REF. 1).
FT	CONFLICT	1763	M -> I (IN REF. 3).
FT	CONFLICT	1788	D -> N (IN REF. 5).
FT	CONFLICT	1871	R -> G (IN REF. 5).
FT	CONFLICT	1882	Q -> R (IN REF. 5).
FT	CONFLICT	1890	MISSING (IN REF. 5).
FT	CONFLICT	1933	MISSING (IN REF. 5).
FT	SEQUENCE	1939 AA; 223689 MM; ECB87E7CEB768B6F CRC64;	
Query Match	8.8%;	Score 236.5; DB 1; Length 1939;	
Best Local Similarity	24.4%;	Pred. No. 0.016;	
Matches 146; Conservative 81; Mismatches 196; Indels 175; Gaps 25;			
QY	12	EGAQEPSPQAFAVEABGPGSQAPRKPEGAQARTAQSGALRDVSEELSRQLEDILSTYC 71	
DB	1037	EGSLQEKKVRMDLE-----RAKKLEGGDLKTQES--IMDLND-KLQEEKKKKKE 1086	
QY	72	VNNNGGPGEDGAQGPAPEDA-----EKSRTVVAR-----NGEPPTPVVVGKSP 119	
DB	1087	FDINQ-----QNSKIDEQALALQKKLKNQARLEEELEEAERTAPAKVEKL 1138	
QY	120	SKGDPNTEIROS-DRVG-----DRHRRPOEKKKAGKGLKRYITLLMQTLNLTSP 171	
DB	1139	SDLSRELEISERLEBAGGATSVQIEMNKKRAEFQMRDLFEATL-----OHEA 1189	
QY	172	KLALCKKY-----AELLBEHRNSQMKLLQKKSQOLVOEKDHLARGEHSAVLARSKL 227	
DB	1190	TAAALRKGHADSVAEIGEQIDNLRQVKQLEKESEFKLEDDVTNMMEOIIXAKANLEK 1249	
QY	228	LCH-----ELQNRNRLKEBVGVAR----- 248	
DB	1250	VSTLEDQANEYRVKLEEAQRSLNDFTTQAKLQFENGELARLEKEALISQTRGKLS 1309	
QY	249	-----BEESKREVTSH-FQVTLNDIOLQMEQHNR-----NSK 281	
DB	1310	YTOQMEDKRLQEEEGKAKNALAHQARSARHDCDLLREQYBETEAKELQRVLSKANSE 1369	
QY	282	LR-----QENNELAERLKKLIEQVLEFRSHIDKY-----PMHK--DLQ 318	
DB	1370	VAQWRKYETDARTEELLEAAKKLQRLQDABEABVAVNAKCSSLEKTKHRLQNEIED 1429	
QY	319	QLYDAXLQQQAEQMKABERHQBKDFLKEAVESQRMCELMKQOETHLKQALALYATEK 378	
DB	1430	LMTYDVERSA-----AAALDKQRFKILAE--WKQKYBESQSELESQKZARSSTELF 1484	
QY	379	EFPNTLSKSESVFTFKQEMERWTKI-----KKLEKETMYR 417	
DB	1485	-KUNAYBESLEHETFKRNTNLQBEISDLTQOLGSGGKNVHELEKVRQOLEVKELEQ 1543	
QY	418	SRWESSNKALLEWAEFTKTVDRKBLEGLQVKIQLEKLC-----RALQTERNDLAKRVQD 471	
DB	1544	SALSEA-EASLEHEEGKILR-AQLEFNQIKAEIERKLAEDKDEMEQAKRN--HORVVD 1597	

RA Kutateladze T.G., Ogburn K.D., Watson W.T., de Beer T., Emr S.D.,
RA Burd C.G., Overduin M.;
RT "Phosphatidylinositol 3-phosphate recognition by the FYVE domain.";
RL Mol. Cell 3:805-811(1999).
RN [7]
RP MUTAGENESIS OF TRP-1349; CYS-1358; PHE-1365; ARG-1370; ARG-1371;
RP HIS-1372; HIS-1373; CYS-1374; ARG-1375; CYS-1377; GLY-1378; CYS-1385;
RP ARG-1400 AND CYS-1405, SUBCELLULAR LOCATION, AND INTERACTION WITH
RP PHOSPHATIDYLINOSITOL 3-PHOSPHATE.
RX MEDLINE=20387352; PubMed=10807926;
RA Gaullier J.-M., Roennigk E., Gillooly D.J., Stenmark H.;
RT "Interaction of the EEA1 FYVE finger with phosphatidylinositol
RT 3-phosphate and early endosomes. Role of conserved residues.";
RL J. Biol. Chem. 275:24595-24600(2000).
RN [8]
RN INTERACTION WITH RAB22A.
RX MEDLINE=21859373; PubMed=11870209;
RA Kauppi M., Simonsen A., Bremnes B., Vieira A., Callaghan J.M.,
RA Stenmark H., Oikarinen V.M.;
RT "The small GTPase Rab22 interacts with EEA1 and controls endosomal
RT membrane trafficking.";
RL J. Cell Sci. 115:899-911(2002).
RN [9]
RN MUTAGENESIS OF GLU-39; PHE-41; ILE-42; PRO-44; MET-47 AND TYR-60,
RP HOMODIMERIZATION, AND INTERACTION WITH RAB5C.
RX MEDLINE=22495953; PubMed=12493736;
RA Merithew E., Stone C., Bathiraj S., Lambright D.G.;
RT "Determinants of Rabs' interaction with the N terminus of early
RT endosome antigen 1.";
RL J. Biol. Chem. 278:8494-8500(2003).
RN [10]
RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1289-1411 IN COMPLEX WITH
RP PHOSPHATIDYLINOSITOL 3-PHOSPHATE, AND HOMODIMERIZATION.
RX MEDLINE=21617582; PubMed=11741531;
RA Dumas J.J., Merithew E., Sudharshan E., Rajamani D., Hayes S.,
RA Lawe D., Corvera S., Lambright D.G.;
RT "Multivalent endosome targeting by homodimeric EEA1.";
RL Mol. Cell 8:947-958(2001).
RN [11]
RN STRUCTURE BY NMR OF 1346-1410 ALONE AND IN COMPLEX WITH
RP PHOSPHATIDYLINOSITOL 3-PHOSPHATE.
RX MEDLINE=21143489; PubMed=11230696;
RA Kutateladze T.G., Overduin M.;
RT "Structural mechanism of endosome docking by the FYVE domain.";
RL Science 291:1793-1796(2001).
CC -I- FUNCTION: Binds phospholipid vesicles containing
CC phosphatidylinositol 3-phosphate and participates in endosomal
CC trafficking.
CC -I- SUBUNIT: Homodimer. Binds STX6. Binds RAB5A, RAB5B, RAB5C and
CC RAB22A that have been activated by GTP-binding.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein
CC associated with early endosomes.
CC -I- DOMAIN: The FYVE-type zinc finger domain mediates interactions
CC with phosphatidylinositol 3-phosphate.
CC -I- DISEASE: Antibodies against EEA1 are found in sera from patients
CC with subacute cutaneous lupus erythematosus and other autoimmune
CC diseases.
CC -I- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC -I- SIMILARITY: Contains 1 FYVE-type zinc finger.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; L40157; AAA79121.1; -;
DR EMBL; X78998; CAAS5632.1; -;
DR PIR; A57013; A57013.
DR PDB; 1HYJ; 14-MAR-01.
DR PDB; 1HYJ; 14-MAR-01.

DR PDB; 1JOC; 28-DEC-01..
DR Genew; HGNC:3185; EEA1.
DR MIM; 605070; -.
DR GO; GO:0005737; C:cytoplasm; TAS.
DR GO; GO:0005769; C:early endosome; TAS.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0005545; P:phosphatidylinositol binding; TAS.
DR GO; GO:0006899; P:nonspecific vesicle transport; TAS.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR000306; Znf_FYVE.
DR Pfam; PF01363; FYVE; 1.
DR Pfam; PF00096; zf-C2H2; 1.
DR SMART; SM00064; FYVE; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS0178; ZF_FYVE; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_2; 1.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
KW Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
FT DOMAIN 74 1348
FT COILED COIL (POTENTIAL).
FT DOMAIN 97 758
FT GLN/GLU/LYS-RICH.
FT DOMAIN 937 1032
FT GLN/GLU/LYS-RICH.
FT DOMAIN 1093 1231
FT C2H2-TYPE.
FT ZN_FING 41 64
FT ZN_FING 1352 1410
FT MUTAGEN 39 39
FT E-->A: STRONGLY REDUCES INTERACTION WITH
FT RAB5C.
FT F-->A: STRONGLY REDUCES INTERACTION WITH
FT RAB5C.
FT I-->A: STRONGLY REDUCES INTERACTION WITH
FT RAB5C.
FT P-->A: STRONGLY REDUCES INTERACTION WITH
FT RAB5C.
FT M-->A: STRONGLY REDUCES INTERACTION WITH
FT RAB5C.
FT Y-->A: STRONGLY REDUCES INTERACTION WITH
FT RAB5C.
FT W-->A: REDUCES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT D-->V: REDUCES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT N-->D: REDUCES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT C-->S: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT P-->A: STRONGLY REDUCES
FT PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING
FT AND ENDOSOMAL LOCATION.
FT VT-->EE,GG: ABOLISHES
FT PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING
FT AND ENDOSOMAL LOCATION.
FT R-->A: ABOLISHES ENDOSOMAL LOCATION.
FT R-->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT H-->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT C-->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT R-->G: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT C-->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT G-->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT C-->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT R-->G: STRONGLY REDUCES
FT PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING
FT AND ABOLISHES ENDOSOMAL LOCATION.
FT C-->S: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT C-->S (IN REF. 1).
FT LQ --> FE (IN REF. 1).

"A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation.";
 Cell 62:999-1006(1990).
 [13]
 VARIANTS CMH1 GLN-249; GLN-403; CYS-453 AND MET-606.
 MEDLINE=92204193; PubMed=1552912;
 Watkins H., Rosenzweig A., Hwang D.S., Levi T., McKenna W., Seidman C.E., Seidman J.G.;
 "Characteristics and prognostic implications of myosin missense mutations in familial hypertrophic cardiomyopathy.";
 New Engl. J. Med. 326:1108-1114(1992).
 [14]
 VARIANTS CMH1 GLN-403; CYS-453; ARG-584 AND MET-606.
 MEDLINE=94070863; PubMed=8250038;
 Watkins H., Thierfelder L., Anan R., Jarcho J., Matsumori A., McKenna W., Seidman J.G., Seidman C.E.;
 "Independent origin of identical beta cardiac myosin heavy-chain mutations in hypertrophic cardiomyopathy.";
 Am. J. Hum. Genet. 53:1180-1185(1993).
 [15]
 VARIANTS CMH1 GLN-403 AND VAL-908.
 MEDLINE=92346810; PubMed=1638703;
 Epstein N.D., Cohn G.M., Cyran F., Pananapazir L.;
 "Differences in clinical expression of hypertrophic cardiomyopathy associated with two distinct mutations in the beta-myosin heavy chain gene. A 908Leu-->Val mutation and a 403Arg-->Gln mutation.";
 Circulation 86:345-352(1992).
 [16]
 VARIANTS CMH1 LEU-403 AND TRP-403.
 MEDLINE=94075629; PubMed=8254035;
 Dausse E., Komajda M., Feller L., Dubourg O., Dufour C., Carrier L., Wisniewsky C., Bercovic J., Hengstenberg C., Al-Mahdawi S.;
 "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and identification of a hot spot for mutations in the beta-myosin heavy chain gene.";
 J. Clin. Invest. 92:2807-2813(1993).
 [17]
 VARIANT CMH1 TRP-403.
 MEDLINE=94093568; PubMed=8268932;
 Moolman J.C., Brink P.A., Corfield V.A.;
 "Identification of a new missense mutation at Arg403, a Cpg mutation hotspot, in exon 13 of the beta-myosin heavy chain gene in hypertrophic cardiomyopathy.";
 Hum. Mol. Genet. 2:1731-1732(1993).
 [18]
 VARIANT CMH1 ASN-615.
 MEDLINE=93038688; PubMed=1417858;
 Nishi H., Kimura A., Harada H., Toshima H., Sasazuki T.;
 "Novel missense mutation in cardiac beta myosin heavy chain gene found in a Japanese patient with hypertrophic cardiomyopathy.";
 Biochem. Biophys. Res. Commun. 188:379-387(1992).
 [19]
 VARIANT CMH1 GLY-778.
 MEDLINE=93343938; PubMed=8343162;
 Harada H., Kimura A., Nishi H., Sasazuki T., Toshima H.;
 "A missense mutation of cardiac beta-myosin heavy chain gene linked to familial hypertrophic cardiomyopathy in affected Japanese families.";
 Biochem. Biophys. Res. Commun. 194:791-798(1993).
 [20]
 VARIANT CMH1 VAL-908.
 MEDLINE=93168485; PubMed=8435239;
 Al-Mahdawi S., Chamberlain S., Cleland J., Nihoyannopoulos P., Gilligan D., French J., Choudhury L., Williamson R., Oakley C.;
 "Identification of a mutation in the beta cardiac myosin heavy chain gene in a family with hypertrophic cardiomyopathy.";
 Br. Heart J. 69:136-141(1993).
 [21]
 VARIANT CMH1 TRP-719.
 MEDLINE=95179132; PubMed=7874131;
 Greve G., Bachinski L., Friedman D.L., Czernuszewicz G., Anan R., Towbin J., Seidman C.E., Roberts R.;
 "Isolation of a de novo mutant myocardial beta MHC protein in a

pedigree with hypertrophic cardiomyopathy.";
 Hum. Mol. Genet. 3:2073-2075(1994).
 [22]
 VARIANTS CMH1 CYS-513; ARG-716 AND TRP-719.
 MEDLINE=94110336; PubMed=8282798;
 Anan R., Greve G., Thierfelder L., Watkins H., McKenna W., Solomon S., Vecchio C., Shono H., Nakao S., Tanaka H., Mares A. Jr., Towbin J.A., Spirito P., Roberts R., Seidman J.G., Seidman C.E.;
 "Prognostic implications of novel beta cardiac myosin heavy chain gene mutations that cause familial hypertrophic cardiomyopathy.";
 J. Clin. Invest. 93:280-285(1994).
 [23]
 VARIANT CMH1 THR-797.
 MEDLINE=96047159; PubMed=7501410;
 Moolman J.C., Brink P.A., Corfield V.A.;
 "Identification of a novel Ala797Thr mutation in exon 21 of the beta-myosin heavy chain gene in hypertrophic cardiomyopathy.";
 Hum. Mutat. 6:197-198(1995).
 [24]
 VARIANT CMH1 CYS-453.
 MEDLINE=96209901; PubMed=8655135;
 Ko Y.-L., Chen J.-J., Tang T.-K., Cheng J.-J., Lin S.-Y., Liou Y.-C., Kuan P., Wu C.-W., Lien W.-P., Liew C.-C.;
 "Malignant familial hypertrophic cardiomyopathy in a family with a 453Arg-->Cys mutation in the beta-myosin heavy chain gene: coexistence of sudden death and end-stage heart failure.";
 Hum. Genet. 97:585-590(1996).
 [25]
 VARIANTS CMH1 THR-349 AND TRP-719.
 MEDLINE=98204402; PubMed=9544842;
 Jeschke B., Uhl K., Weist B., Schroder D., Meitinger T., Dohlemann C., Vosberg H.-P.;
 "A high risk phenotype of hypertrophic cardiomyopathy associated with Query Match 8.7%; Score 234; DB 1; Length 1935;
 Best Local Similarity 22.4%; Pred. No. 0.019;
 Matches 127; Conservative 100; Mismatches 177; Indels 162; Gaps 25;
 QY 8 EAGPEGAQERPSQAAPAV-----EAGPGSSQAPRKEGAQARTAQSCALRDVSEELS 60
 DB 1363 KANSEVAQWRKYVETDAIORTBELEAKKLAQRLQEAEEAVEAVNAKCSLEKTKRLQ 1422
 QY 61 RQEDILSTYCVDDNQGPGEDGAGCPAPEDDAEKSRITYVARNGEPEPTFVYGEKFS 120
 DB 1423 NEIEDLM-----VDVERSNAAL-----DKQOR 1447
 QY 121 KGPNTETIROSDVGRDHRPQEKKAKGLKGTITLLM-----OTLNTLSTPEEKLAAL 176
 DB 1448 NFDKILAEWKQYVESQSELESSQ--KEARSLSTELFKLNAYESLEHLETFKRENKL 1505
 QY 177 CKYAEALLBEHRNSOKMKLQKQSQVLQVEKDLRG-----EHSKAVLARSKLE- 226
 DB 1506 QBEISDLTEQLGSSGKTIHELEKVRKQLEAKRMELQSALEAEASLEHEEGKILRAQLBF 1565
 QY 227 -----SICRELQRHNSL-----KEEG----- 243
 DB 1566 NQKAFIEKLAEKDEMEQAKNNLRVDSLSQTDATRRNRNEALRVKKGEGDLNEM 1625
 QY 244 -VQRA---REBEKKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMEIAR----- 292
 DB 1626 EIQLSHANEMAAEAQKVKS-LQSLKQDTQIQLDDAVRANDDLK-ENIAIVERRNNLQA 1683
 QY 293 -----LKLIEQYE-----LREEHIDKVFKEKDI-----QQQLVDKLAQAEMLKE 334
 DB 1684 ELEEELRAVVEQTERSKRLABQSLIETSERVQLHSHQNTSLNQKQODADLSQLQTEVESE 1743
 QY 335 A--EERHOREKDFLLKEAVESQRM--BLMKQOET--HLKQOLALYTEFESFQNTLSKS 388
 DB 1744 AVQECENAEK---AKKAITDAAMAEELKKEQDTSALHRMKONMEQITIKDQRLDEA 1800
 QY 389 SEV-FTTFQEKERKMTKKIKLEKETTVYRSRWSSNKALLEAEKTVRDEK-LEGLOV 446
 DB 1801 EQIALGKGKKQLQKLEARVRELENE-----LE-AEQK--RNAESVKGMRK 1842

Qy 447 KIQRLKLCALQTERNDLNKRVQDL 472
Db 1843 SERRIKELTYQTEBDRKNL-LRLQDL 1867

RESULT 30

MYH7 FIG STANDARD; PRT; 1935 AA.
AC P79293;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).
GN MYH7.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Domestica;
RA Ko Y.L.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBSJ databases.
CC -!- FUNCTION: Muscle contraction.
CC -!- SURUNIT: Muscle myosin is a hexameric protein that consists of 2 heavy chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2 regulatory light chain subunits (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing cycles of a 28-residue repeat pattern composed of 4 heptapeptides, characteristic for alpha-helical coiled coils.
CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be split further into 2 globular subfragments (S1) and 1 rod-shaped subfragment (S2).
CC -!- MISCELLANEOUS: The cardiac alpha isoform is a 'fast' ATPase myosin, while the beta isoform is a 'slow' ATPase.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; U75316; AAS37320.1; -.
CC PIR; A59286; A59286.
CC HSP; P08799; 1MND.
CC InterPro; IPR000048; IQ region.
CC InterPro; IPR01609; myosin head.
CC InterPro; IPR004009; Myosin_N.
CC InterPro; IPR002828; Myosin_tail.
CC Pfam; PF00612; IQ; 2.
CC Pfam; PF00063; myosin head; 1.
CC Pfam; PF02736; Myosin_N; 1.
CC Pfam; PF01576; Myosin_tail; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00242; MSc; 1.
CC PROSITE; PFS0096; IQ; 1.
CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Multigene family; Calmodulin-binding.
FT DOMAIN 1 780
FT MYOSIN HEAD-LIKE.
FT DOMAIN 781 810
FT COILED COIL (POTENTIAL).
FT DOMAIN 839 1935
FT ATP (POTENTIAL).
FT NP_BIND 178 185
FT ACTIN-BINDING.
FT DOMAIN 655 677
FT ACTIN-BINDING.
FT DOMAIN 757 771

FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
SQ SEQUENCE 1935 AA; 223110 MW; 149CDBF910DBB08 CRC64;
Query Match 8.7%; Score 234; DB 1; Length 1935;
Best Local Similarity 22.3%; Pred. No. 0.019;
Matches 126; Conservative 101; Mismatches 177; Indels 162; Gaps 25;
Qy 8 EAGPGGAQEPSPQAPAV-----EAGPGSSQAPKDEGACARTAQSGALRDVSEELS 60
Db 1363 KANSEVAQWRKYETDAIORTLEELAKKKLQADAEVAEVAIVAKSSLEKTKRLQ 1422
Qy 61 RQLEDILSTYCVDMNQGPEDGAGQEPAPDAKSRITYVARNGEPEPTPVVYGRKEPS 120
Db 1423 NEIEDLM-----VDVERSNAAAAAL-----DKOR 1447
Qy 121 KGPNTETIRQSDVGRDHRHROEKKKAGLKEITLM-----QTLNTLSTPEKLAAL 176
Db 1448 NFDKILAEWKQYEEQSLESSQ--KEARSLSTFLKLNAYEESLEHLETSKRKNKL 1505
Qy 177 CKVAELLBEHRNSQOMKLLQKQSQLVQEKDHLFG-----EHSKAVLARSKE- 226
Db 1506 QEBISDLTEQLGSSGKTIHELEKVKQLAEKLEQLQSALEESSEASLEHEEGKILRAQLEF 1565
Qy 227 -----SLCRLQHNRL-----KEEG----- 243
Db 1566 NQIKAEERKLAEBEQAENHLRVVDSLQTSLSDAETRSNEALRVKKQMEGDLNEM 1625
Qy 244 -VQRA--REBEKKREVTSHFOVTINDIQLOMEQHNERNKSLROENMELAE- 292
Db 1626 SIQSHANNWAAQAQKQVKS-LQSLKDTQIQDLDVANDDLK-ENIAIVERNNLQA 1683
Qy 293 ----LKLKIQYE----LREHIDKVFVKHDL-----QQQLVDAKLOQAQEMKE 334
Db 1684 ELBELRAVVEQTERSRLAEQELIETSERVQLLSQNTSLINQKKMEADLSQLEVEE 1743
Qy 335 A--EESHOREKDFLLKEAVESQWC--ELMKQOET--HLKQALALYKEPEEFQNTLSKS 388
Db 1744 AVQESRNAEEK---AKKAITDAAMAEHLKQEQUTSAHLERKKNMEQTIKIQHRLDEA 1800
Qy 389 SEV-FTTFQEMEKMTYKIKKLEKETTMYRSRVSSNKALLEMAEKTVDKE-LEGLOV 446
Db 1801 EQIALGKQKQQLKLEARVRELENE-----LE-AEQK--RNAESVKGMRK 1842
RESULT 31
SCPI_MESAU STANDARD; PRT; 845 AA.
AC Q60563;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Synaptonemal complex protein 1 (SCP-1 protein) (Meiotic chromosome synaptonemal complex protein) (Fragment).
DE SCPI OR SYN1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=95181577; PubMed=7876343;
RA Dobson M.J., Pearlman R.E., Karaiskakis A., Spyropoulos B., Moens P.B.;
RA "Synaptonemal complex proteins: occurrence, epitope mapping and chromosome disjunction."
RL J. Cell Sci. 107:2749-2760(1994).
CC -!- FUNCTION: Major component of the transverse filaments of

```
CC synaptonemal complexes (SCS), formed between homologous
CC chromosomes during meiotic prophase (By similarity). Has non-
CC specific DNA binding capability.
CC -|- SUBCELLULAR LOCATION: Nuclear. In tripartite segments of
CC synaptonemal complexes, between lateral elements in the nucleus.
CC Found only where the chromosome cores are synapsed. Its N-terminus
CC is found towards the centre of the synaptonemal complex while the
CC C-terminus extends well into the lateral domain of the
CC synaptonemal complex.
CC -|- DOMAIN: Consists of an alpha-helical stretch of 700 AA residues,
CC flanked by N- and C-terminal globular domains. The C-terminal
CC domain has DNA-binding capacity (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L32978; AAC42039.1; -.
CC PIR; I48176; I48176.
CC InterPro; IPR008827; SCP-1.
CC Pfam; PF05483; SCP-1; 1.
CC KW Nuclear protein; Melosis; Cell division; Phosphorylation;
CC DNA-binding; Coiled coil.
CC NON TER 1
CC DOMAIN <1 672 COILED COIL (POTENTIAL).
CC DOMAIN 553 556 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC DOMAIN 753 756 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC DOMAIN 830 838 ARG/LYS-RICH (BASIC).
CC SEQUENCE 845 AA; 99401 MW; D7F28873C824C6A8 CRC64;
CC -----
Query Match 8.6%; Score 233.5; DB 1; Length 845;
Best Local Similarity 24.0%; Pred. No. 0.0091;
Matches 109; Conservative 86; Mismatches 168; Indels 91; Gaps 19;
Qy 90 EPEDAKSRVTYVANGPEPTPVYVYGEKSPKSGDPNTEETIRQSDVCGDRP----- 139
Db 54 EYEEETRVQVYVYVNNNIEMILAF-----BELRVQVARNALDMHFKLEDHE 101
Qy 140 ---HRRPQEKKAGLGEITFLAQTLNTLSTPEKLAALCKYAELEEHRSQKQ-- 194
Db 102 KIQLQBEYKKEVNDKENQVSL--LIQRTKENKM---KDLTFLESDRKNVQLED 154
Qy 195 ---KLQKQKQSVQKQHLRG--BHSKAVLARS--KLESICRELQKRNRLSKKEGVQAR 248
Db 155 KTKLQDENVKELNKKKHILTSLEDTKMSLQSRSMNTQKALEEDLIQATKTIYQUTEKEA 214
Qy 249 BEERKKEVTSF--FQV-----TLNDIOLQMEQHN--ERNSKLRQENMELAEKLLI 297
Db 215 QMEFPNKAQTHDSFMTVKATTCITLBEI-LRTQQRVLVKNEDQLKILTWELQKSNELD 273
Qy 298 EYELRBEHDKVPKHKDL---QQQLVDKIQQAQEMLKAEERHQ-----REKDF- 345
Db 274 EMTKFNQNNKVELKELKILAEQDKLDEK-KQVEKLAELQGEQBELTLLQTREREVH 332
Qy 346 -----LKAEVSRKMCLEKQVTHLQKQ-----LALYTEKFEFQFQ-- 383
Db 333 DLERQLLVTKISDQNYSKQVSELKTKGEEKKNKAEITASCCKSLNNKTKQTNDMAL 392
Qy 384 TILSKSSVFYTFKQEMENQTKIKLEKETTWYSRWNSSNKALLEAE-RTVDRKELE 442
Db 393 ELAKYQEDITNSKQEEERMLQIENLEBKETHLRLDELESVKEFTIQGNEVKCKLDKSEE 452
Qy 443 ---GLQVKIQLEKLCRALQTERNDLNKRVQDLS 473
Db 453 NARSIECEVLKQKQMKILNKKNLNRKQAEKNS 486
RESULT 32
MYH2_HUMAN
```

```
ID MYH2_HUMAN STANDARD; PRT; 1941 AA.
AC Q9UXK2; Q14322; Q16229;
DC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, skeletal muscle, adult 2 (Myosin heavy chain Ila)
DE (MyHC-IIa).
GN MYH2 OR MYHSA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=9918869; PubMed=10388558;
RA Weiss A., Schiaffino S., Leinwand L.A.;
RT "Comparative sequence analysis of the complete human sarcomeric myosin
RT heavy chain family: implications for functional diversity.";
RL J. Mol. Biol. 290:61-75(1999).
RN [2]
SEQUENCE OF 1711-1941 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=95109625; PubMed=7545970;
RA Smerdu V., Karsch-Mizrachi I., Campione M., Leinwand L.,
RA Schiaffino S.;
RT "Type Iix myosin heavy chain transcripts are expressed in type Iib
RT fibers of human skeletal muscle.";
RL Am. J. Physiol. 267:C1723-C1728(1994).
RN [3]
SEQUENCE OF 1823-1941 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=95270723; PubMed=75151403;
RA Emmon S., Sant'ana Pereira J., Sargeant T., Young A., Goldspink G.;
RT "Characterization of human skeletal muscle fibres according to the
RT myosin heavy chains they express.";
RL J. Muscle Res. Cell Motil. 16:35-43(1995).
CC -|- FUNCTION: Muscle contraction.
CC -|- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC and 2 regulatory light chain subunits of the myofibrils.
CC -|- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -|- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -|- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be
CC split further into 2 globular subfragments (S1) and 1 rod-shaped
CC subfragment (S2).
CC -|- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -|- SIMILARITY: Contains 1 IQ domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF111784; AAD29950.1; -.
CC EMBL; S73840; AAC33916.1; -.
CC EMBL; Z32858; CAAB3687.1; -.
CC PIR; I51912; I51912.
CC HSSP; P13538; 2MYS.
CC Genew; HGNC:7572; MYH2.
CC MTM; 160740; -.
CC GO; GO:0005859; C:muscle myosin; TAS.
CC GO; GO:0003779; P:actin binding; NAS.
CC GO; GO:0005524; P:ATP binding; NAS.
CC GO; GO:0005516; P:calmodulin binding; NAS.
CC GO; GO:0003776; P:muscle motor activity; TAS.
CC GO; GO:0006936; P:muscle contraction; TAS.
```


DR Pfam; PF00063; myosin head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSIN_HEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding; ATP-binding; Methylation; Multigene family; Calmodulin-binding.
KW ATP-binding; Methylation; Multigene family; Calmodulin-binding.
FT DOMAIN 1 780
FT DOMAIN 781 810 IQ
FT DOMAIN 840 1935
FT NP_BIND 178 185 COILED COIL (POTENTIAL).
FT DOMAIN 655 677 ATP.
FT DOMAIN 757 771 ACTIN-BINDING.
FT MOD_RES 129 129 ACTIN-BINDING.
FT CONFLICT 1529 1531 METHYLATION (TRI-) (POTENTIAL).
FT CONFLICT 1731 1731 IRK -> VRR (IN REF. 3).
FT CONFLICT 1784 1784 D -> H (IN REF. 3).
FT CONFLICT 1784 1784 N -> K (IN REF. 3).
FT CONFLICT 1851 1851 T -> N (IN REF. 3).
FT CONFLICT 1851 1851 R -> K (IN REF. 3).
FT CONFLICT 1858 1858
SQ SEQUENCE 1935 AA; 223082 MW; C8376C324A7BD82B CRC64;

Query Match 8.6%; Score 232; DB 1; Length 1935;
Best Local Similarity 22.3%; Pred. No. 0.023;
Matches 116; Conservative 81; Mismatches 189; Indels 134; Gaps 16;

QY 25 VEAEGPGSQAPKPEGAQTAQSGALRDVSELSQLEDILSTVCVNNQGGPGDGA 84
DB 1094 IEDEQALGSLQKXLEQARI-----BELSELE----- 1123

QY 85 QGPAPPEDEKERTVYVARGPEPTPVVYGEKPSKGDENTBEIRQSDVGVDRHRRPQ 144
DB 1124 -AERTAKVKEKRSLSRELEB-----ISERLEAGGATSVQI-----EMKKREA 1169

QY 145 EKKAGKGLGKEITLMTLNTLSTPEKLAALCKKY-----AELEHRSQKQKLLQKK 200
DB 1170 EFQKRRDLSEATL-----QHEATAAALKKHADSVAEISGEQIDMLQRYKQKLEKE 1220

QY 201 QSQVQEKQHLREHSHKAVLARSKLSLCT-----EL 232
DB 1221 KSEFKLEDDVTSNMEQIIKAKANLEKCKCTLSDQNEHRSKABETQVSNDLTQRAKL 1280

QY 233 QRHN-----RSLKE-----EGVQARBEERKKEVTSN-FQVTLNDI 268
DB 1281 QTENGELSRQDEKEALISQITRGKLTQYQLEDLKRQLEEEVKAKNALAHQSAHDHC 1340

QY 269 QLOMEQHNSK-----LRQNNVELARLKKLEIQYELREHIDKVKFKHQLQQQLV 321
DB 1341 DLLREQYEETRAKAEQLQRYLSKANSEVAQWRKTYETDAIQTTELEEA--KKGLAQRLQ 1398

QY 322 DAKLQQAQE-----MLKEAEERHOREKDFLKEAVESQRMCELMKQQTTHLKQQLALY 374
DB 1399 DA--EAEVAVNAKCSSEKTEKHLQNEIEDLVAVDVSNAANAALDKKQNFQKILVEN 1456

QY 375 TEKPEPONTLSKSEVFTTFKQMEKQTKKIKLEKFTNYSRVSSESSKALLEMAEK 434
DB 1457 KQYVESQSESSQKARSLSLTEL-----FLKNAVSESLHETFKRENKQLQEBISDL 1512

QY 435 TVRDKEGLQVKIQRLKCLRALOTERNLNKRKQDLSA 474
DB 1513 T---EQLGSTGKSIHELEKTKQLEAEKLELSQSALEAEAA 1549

RESULT 34
RA50_PVRFU
ID RA50_PVRFU STANDARD; PRT; 882 AA.
AC P5B301;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.

GN RAD50 OR PFI167.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]_TaxID=2261;
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX MEDLINE=20485553; PubMed=11029422;
RA Hopfner K.-P., Karcher A., Shin D., Fairley C., Tainer J.A.,
Carney J.P.,
RT "Mrell and Rad50 from Pyrococcus furiosus: cloning and biochemical
characterization reveal an evolutionarily conserved multi-protein
machine";
RT J. Bacteriol. 182:6036-6041(2000).
RL [2]
RN SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 1-149.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX Hopfner K.-P., Karcher A., Shin D.S., Craig L., Arthur L.M.,
Carney J.P., Tainer J.A.;
RT "Structural biology of Rad50 ATPase: ATP-driven conformational
control in DNA double-strand break repair and the ABC-ATPase
superfamily";
RT Cell 101:789-800(2000).
RL CC
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
rad50/mrell complex possesses single-strand endonuclease activity
and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding
and/or repositioning DNA ends into the mrell active site.
CC -!- SUBUNIT: Forms a complex with mrell.
CC -!- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AB010225; AAL81291.1; -.
DR PDB; 1P2T; 20-SEP-00.
DR PDB; 1P2U; 02-AUG-00.
DR PDB; 1I18; 30-MAY-01.
DR PDB; 1L8D; 28-AUG-02.
DR HAMAP; MF_00449; -; 1.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR007517; Rad50 zn_hook.
DR Pfam; PF04423; Rad50 zn_hook; 1.
DR Pfam; PF02463; SMC_N; 1.
DR ProDom; PD000006; ABC transporter; 1.
KW DNA repair; Hydrolyase; ATP-binding; Coiled coil; 3D-structure;
KW Complete proteome.
FT NP_BIND 30 37 ATP.
FT DOMAIN 148 744 COILED COIL (POTENTIAL).
SQ SEQUENCE 882 AA; 103839 MW; 3ADCB250382A99E CRC64;

Query Match 8.6%; Score 231.5; DB 1; Length 882;
Best Local Similarity 23.3%; Pred. No. 0.011;
Matches 103; Conservative 90; Mismatches 148; Indels 101; Gaps 14;

QY 127 EBIROSDEVGDRDHRPQEKKAGLKGKTEITLMTLNTLSTPEEKALCKYAELEE 186
DB 144 DAILESDEAREKVRVLEVLNLDKFEAYKLSLKKTTNNRIKEYRDLARTENIELKE 203

QY 187 HRSOKOMKLOK-----KQSQ-L-VQSKDHRLGHEHSAVLARSK----- 224

Db 204 --NEQLIQVQISKIEVLSKSKVDMLEKKEVLRLEETKVEIENSERLEKRGDKR 261

QY 225 -LESCLQRNRSLKRGVQARAEKKEKREKVS----- 259

Db 262 TLPERIKNTAEVLEKKE-----KEKELEQVKEITSIKKQVDAYLALKEPKNEVLDKKYK 317

QY 260 -----HFQVTLNDIQLOEQEHNERS-----KLQENQWELAEKLLIQEYELREHIDKV 310

Db 318 IEKELTRVELINIQRIEELNEKSEKKEKNEKKEKILNKLALDKHQYEEIKAKK 377

QY 311 FKHDIQOQVDAKQQAQEMLKE-----AEERHQ-----RKQFLLEAVE-- 352

Db 378 ENLRQLKELGDKSPEDIKKLELELETKKTTEERNEITQRIQELKKNKIGDLKTAIBEL 437

QY 353 --SORMCELMKQOET--HLKQOLALVTEKPEFQNTLSKSEVFTTFKQEMKTKKIKK 408

Db 438 KKAQKQKPCVGRBLTDEHRELLSKYHLDNNSKNTLAKLDRKSELERELRTRDMEIKR 497

QY 409 LEKETTM--YRSRWESSNKALLEMAEK-TYRDKELGLOV-----KIQRL 451

Db 498 LTPLLTVAEQIRSIIEELNVNLEKIEKNATEYKELLBELRTLEGRIRGLAEDLKLAFL 557

QY 452 EKLCRALQOTERNDLNKRVODLS 473

Db 558 EKXLAALIHKKQELKELKELN 579

RESULT 35

MYH6 MOUSE

AC Q02556; Q64258; Q64738; PRT; 1938 AA.

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-2004 (Rel. 43, Last annotation update)

DE MYOSIN heavy chain, cardiac muscle alpha isoform (MyHC-alpha).

GN MYH6 OR MYHCA.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RQ SEQUENCE FROM N.A.

RC STRAIN=BALB/c, A/J, C57BL/6J, and DBA/2J;

RX MEDLINE=92250040; PubMed=1577481;

RA Quinn-Laquer B.K., Kennedy J.E., Wei S.J., Beisel K.W.;

RT "Characterization of the allelic differences in the mouse cardiac

RT alpha-myosin heavy chain coding sequence.";

RL Genomics 13:176-188(1992).

RN [2]

RQ SEQUENCE OF 1-67 FROM N.A.

RC STRAIN=AKR;

RX MEDLINE=91225025; PubMed=2026617;

RA Gulick J., Subramaniam A., Neumann J., Robbins J.;

RT "Isolation and characterization of the mouse cardiac myosin heavy

RT chain genes.";

RL J. Biol. Chem. 266:9180-9185(1991).

CC -1- FUNCTION: Muscle contraction.

CC -1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2

CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)

CC and 2 regulatory light chain subunits (MLC-2).

CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.

CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing

CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,

CC characteristic for alpha-helical coiled coils.

CC -1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light

CC meromyosin (LMW) and 1 heavy meromyosin (HMW). It can later be

CC split further into 2 globular subfragments (S1) and 1 rod-shaped

CC subfragment (S2).

CC -1- MISCELLANEOUS: The cardiac alpha isoform is a 'fast' ATPase

CC myosin, while the beta isoform is a 'slow' ATPase.

CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.

CC -1- SIMILARITY: Contains 1 IQ domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC

DR EMBL; M76598; AAA37159.1; -

DR EMBL; M76599; AAA37160.1; -

DR EMBL; M76600; AAA37161.1; -

DR EMBL; M76601; AAA37162.1; -

DR EMBL; M62404; AAA37424.1; -

DR PIR; I49464; I49464.

DR HSP; P08799; LMND.

DR SWISS-2DPAGE; Q02566; MOUSE.

DR MG; MG1:97255; Myh6.

DR InterPro; IPR000048; IQ_region.

DR InterPro; IPR001609; Myosin head.

DR InterPro; IPR004009; Myosin N.

DR InterPro; IPR002928; Myosin_tail.

DR Pfam; PF00612; IQ; 2.

DR Pfam; PF00063; myosin head; 1.

DR Pfam; PF02736; Myosin_N; 1.

DR Pfam; PF01576; Myosin_tail; 1.

DR PRINTS; PR00193; MYOSINHEAVY.

DR PRODOM; PD000355; myosin head; 1.

DR SMART; SM00015; IQ; 1.

DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS50096; IQ; 1.

KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;

KW ATP-binding; Methylation; Multigene family; Calmodulin-binding;

KW Polymorphism.

FT DOMAIN 1 782 MYOSIN HEAD-LIKE.

FT DOMAIN 783 812 IQ.

FT DOMAIN 842 1938 COILED COIL (POTENTIAL).

FT NP_BIND 178 185 ATP.

FT DOMAIN 657 679 ACTIN-BINDING.

FT DOMAIN 759 773 ACTIN-BINDING.

FT DOMAIN 790 807 CALMODULIN-BINDING (BY SIMILARITY).

FT DOMAIN 816 833 CALMODULIN-BINDING (BY SIMILARITY).

FT MOD_RES 129 139 METHYLATION (TRI-) (POTENTIAL).

FT VARIANT 124 134 Y -> D.

FT VARIANT 545 545 S -> A.

FT VARIANT 838 838 I -> S.

SQ SEQUENCE 1938 AA; 223564 MW; EAD789ADA68818FB CRC64;

Query Match 8.6%; Score 231.5; DB 1; Length 1938;

Best Local Similarity 24.2%; Pred. No. 0.024;

Hatches 143; Conservative 81; Mismatches 195; Indels 173; Gaps 24;

QY 12 EGAQRPSQAAPVABRGSGSQAPRKPEGAQAQTAQSGALRDVSELSRQLEDILSTYC 71

Db 1037 EGSLEQKKVRMDLE-----RAKKLEGLKLTQES--IMDLND-KLQLEKLEKKE 1086

QY 72 VNNQGGPCDGAQGSAPRPDA-----EKSRTYVAR-----NGEPEPTPVVGEKXP 119

Db 1087 FDISQ-----QNSKLEDEQALLAQKLEKQNAIELEHELEAERTAKVEKL 1138

QY 120 SKGDPNTEIROS-DEVG-----DRHRRPQEKKAKGLKGLKTTLLMQLTNTLSTPRE 171

Db 1139 SLSLELEISERLEAGGATSVQENMKKGAEEFQKMRDLSEATL-----OHEA 1189

QY 172 KLAALCKKY-----ALLLEHRSQKMKLOKQKQSVQEKDHLRGHSEKAVLARSKLS 227

Db 1190 TAAALEKHADSVARELGEQIDMLQVQKLEKESEFKLEDDVTSNNQEIKAKANLEK 1249

QY 228 LCR-----ELQRNLSLKEEGVQAR----- 248

Db 1250 VSRITLEDQANRYRVKLEAQRSLNDFTTQRAKLTQENGELARQLEKEBALISQUTRGKLS 1309

1036 EGSLEQEKVKRMDS-----BAKRLGDLKLQDS--IMDLEND-KLQLEBKLLKKE 1085
72 VDNNGGPGDGAQGEAPEDA-----EKSTYVAR-----NGEPTTPVYGEKEP 119
1086 FDISQ-----QNSKIDQALALQLOKXENQARIBELFELEAERTAKVEKLR 1137
120 SKDGPNTTEIRQS-DEVG-----DRHRRPPOKKKAKGLGHEIILWLTTLSTPES 171
1138 SOLTRELEETISERLEEAGGATSVOIENKKBREAFQCRDLLEATL-----QHEA 1188
172 KLAALCKKY-----ABELEHERNSQKMLQKQSOVLQEKDHLRGHSHKAVLARSKLES 227
1189 TAAALRXKHADSVABELGEOIDNLRVKQLEKESBEKLELDVDTSHMEQIIKAKANLEK 1248
228 LCR-----ELQRNRSKKBEGVQBAR----- 248
1249 VSRITLEDQANRYKLFQEAQSRSLNFTTQAKLQTEGELARQLSEKALINWLTGKLS 1308
249 -----BEEKRKEVTH-FQVTLNDLQLOQEHNER-----NSK 281
1309 YTOOMEDLKQLEBEGKAKNALAHALQASRDCDLLEQYEBEWEAKAEIQRVLKANS 1368
282 LR-----QENMELAEURLKLEJOEYELREEHIDKY-----FKHK---DLQ 318
1369 VAQWRKTYETDAIQRTSEELAEAKKLAQLQDAEAVEAVNAKCSSLEKTKHRLQNEIED 1428
319 QLVDAKLOQAQEMLEKABERHOREKDFLLKEAVESORMCELMKQOETHLQOALAYTEKF 378
1429 LMDVRSNA-----AAALDKKQKQNFDKILAE--WKQYYESQSELESSQKREASLTSELP 1483
379 BEFQNTLSKSEVFTTFKQEMKNTKKI-----KKLEKETTMYR 417
1484 -KLNAYEESULEHLETTKRENKLNQEBISDLTEQLEGKGNVHELEKIRQLEVEKLELQ 1542
418 SRWESSKALLEMAEKTVDKLEGLQVKIQRLKLC-----RALQTERDNLKRVQD 471
1543 SALFEA-EASLEHEEGKILR-AQLEFNQIKABIERKLAERKDEMEQAKRNHL--RVD 1596

RESULT 37

RRBI_MOUSE
ID -RRBI_MOUSE STANDARD; PRT; 1605 AA.
AC Q99PL5; Q99PK5; Q99PK6; Q99PK7; Q99PK8; Q99PK9; Q99PL0; Q99PL1;
AC Q99PL2; Q99PL3; Q99PL4; Q9CS20;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosome-binding protein 1 (Ribosome receptor protein) (MRP).
GN RRP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS RRP61/P180; RRP47; RRP41; RRP16.8;
RP RRP15A; RRP15B; RRP10; RRP5.4; RRP2; RRP1.8 AND RRP0), AND TISSUE
RP SPECIFICITY.
RC STRAIN=ICR; TISSUE=Embryo;
RX MEDLINE=21100429; PubMed=11167022;
RA Kim Y.-J., Lee M.-C., Kim S.-J., Chun J.-Y.;
RT "Identification and characterization of multiple isoforms of a mouse
RT ribosome receptor."
RL Gene 261:337-344 (2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay E.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton G., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Smalhus D.E.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE OF 736-1605 FROM N.A. (ISOFORM 3).
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinaigawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saio T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saio R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bulc D., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RT Nature 409:685-690 (2001).
CC -!- FUNCTION: Acts as a ribosome receptor and mediates interaction
CC between the ribosome and the endoplasmic reticulum membrane (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Type III membrane protein. Endoplasmic
CC reticulum (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=13;
CC Comment=Additional isoforms seem to exist. Experimental
CC confirmation may be lacking for some isoforms;
CC Name=3;
CC IsoId=Q99PL5-1; Sequence=Displayed;
CC Name=RRP61; Synonyms=pl180;
CC IsoId=Q99PL5-2; Sequence=VSP_003963, VSP_003964;
CC Name=RRP47;
CC IsoId=Q99PL5-3; Sequence=VSP_003959, VSP_003963, VSP_003964;
CC Name=RRP41;
CC IsoId=Q99PL5-4; Sequence=VSP_003961, VSP_003963, VSP_003964;
CC Name=RRP16.8;
CC IsoId=Q99PL5-5; Sequence=VSP_003954, VSP_003962, VSP_003963,
CC VSP_003964;
CC Name=RRP15a;
CC IsoId=Q99PL5-6; Sequence=VSP_003958, VSP_003963, VSP_003964;
CC Name=RRP15b;
CC IsoId=Q99PL5-7; Sequence=VSP_003955, VSP_003963, VSP_003964;
CC Name=RRP10;
CC IsoId=Q99PL5-8; Sequence=VSP_003956, VSP_003963, VSP_003964;
CC Name=RRP5.4;
CC IsoId=Q99PL5-9; Sequence=VSP_003957, VSP_003963, VSP_003964;
CC Name=RRP2;
CC IsoId=Q99PL5-10; Sequence=VSP_003953, VSP_003963, VSP_003964;
CC Name=RRP1.8;
CC IsoId=Q99PL5-11; Sequence=VSP_003951, VSP_003963, VSP_003964;
CC Name=RRP0; Synonyms=RS130;
CC IsoId=Q99PL5-12; Sequence=VSP_003952, VSP_003963, VSP_003964;
CC Name=2;

CC IsoId=099PL5-13; Sequence=VSP_003960, VSP_003963, VSP_003964;
 CC -I- TISSUE SPECIFICITY: Widely expressed.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AF273683; AAKI1963.1; -
 CC DR EMBL; AF273684; AAKI1964.1; -
 CC DR EMBL; AF273685; AAKI1965.1; -
 CC DR EMBL; AF273686; AAKI1966.1; -
 CC DR EMBL; AF273687; AAKI1967.1; -
 CC DR EMBL; AF273688; AAKI1968.1; -
 CC DR EMBL; AF273689; AAKI1969.1; -
 CC DR EMBL; AF273690; AAKI1970.1; -
 CC DR EMBL; AF273691; AAKI1971.1; -
 CC DR EMBL; AF273692; AAKI1972.1; -
 CC DR EMBL; AF273693; AAKI1973.1; -
 CC DR EMBL; BC031452; AAKI1974.1; -
 CC DR EMBL; AK019664; AAKI1975.1; -
 CC DR MGD; MGI:1932395; Rbpl.
 CC GO: GO:0030176; C: integral to endoplasmic reticulum membrane; ISS.
 CC GO: GO:0030376; F: ribosome receptor activity; ISS.
 CC GO: GO:0007165; P: signal transduction; ISS.
 CC InterPro: IPR007794; Rib. recp KP reg.
 CC Pfam: PF05104; Rib. recp KP reg.
 CC Transport; Protein transport; Translocation; Endoplasmic reticulum;
 CC Transmembrane; Repeat; Alternative splicing.
 CC DOMAIN 1 7 LUMENAL (POTENTIAL).
 CC TRANSMEM 8 28 POTENTIAL.
 CC DOMAIN 29 1605 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 196 807 61 X 10 AA APPROXIMATE TANDEM REPEATS OF
 CC N-Q-A-K-A-R-G-A-P
 CC VARSPLIC 177 789 Missing (in isoform RRP1.8).
 CC VARSPLIC 177 809 Missing (in isoform RRP0).
 CC VARSPLIC 196 787 Missing (in isoform RRP2).
 CC VARSPLIC 205 626 Missing (in isoform RRP1.6).
 CC VARSPLIC 205 666 Missing (in isoform RRP1.5b).
 CC VARSPLIC 209 720 Missing (in isoform RRP1.0).
 CC VARSPLIC 209 766 Missing (in isoform RRP5.4).
 CC VARSPLIC 269 730 Missing (in isoform RRP1.5a).
 CC VARSPLIC 309 449 Missing (in isoform RRP4.7).
 CC VARSPLIC 461 601 Missing (in isoform RRP2).
 CC VARSPLIC 466 666 Missing (in isoform RRP4.1).
 CC VARSPLIC 717 738 Missing (in isoform RRP1.6.8).
 CC VARSPLIC 841 842 GP -> VC (in isoform RRP6.1, isoform
 CC RRP1.8, isoform RRP0, isoform RRP2,
 CC isoform RRP1.6.8, isoform RRP1.5b, isoform
 CC RRP1.0, isoform RRP5.4, isoform RRP1.5a,
 CC isoform RRP4.7, isoform RRP4.1, isoform
 CC RRP1.6.8 and isoform 2).
 CC VARSPLIC 843 1605 Missing (in isoform RRP6.1, isoform
 CC RRP1.8, isoform RRP0, isoform RRP2,
 CC isoform RRP1.6.8, isoform RRP1.5b, isoform
 CC RRP1.0, isoform RRP5.4, isoform RRP1.5a,
 CC isoform RRP4.7, isoform RRP4.1, isoform
 CC RRP1.6.8 and isoform 2).
 CC FT

FT RRP1.6 and isoform 2).
 FT /FTId=VSP_003964.
 FT A -> G (IN REF. 1; AAKI1965).
 FT G -> S (IN REF. 1; AAKI1967).
 SQ SEQUENCE 1605 AA; 172880 MW; 05E0808BF10C665 CRC64;
 Query Match 8.5%; Score 231; DB 1; Length 1605;
 Best Local Similarity 20.1%; Pred. No. 0.021;
 Matches 142; Conservative 108; Mismatches 230; Indels 228; Gaps 27;
 QY 1 KSPSPQPEAGPGP-----AQERPSQAPAVEAGPGSSQAPRKP----- 39
 DB 801 KISEMAPAQGQKAMVQSQAPQAPAKKSGSKGPGPPDCDGPFLPKYLVSTV 860
 QY 40 -----EGAAQAR-----TAQSCALRDVSEE-----LSRQL-----EDILSTYCVDMN 75
 DB 861 GSVFSEGEAQLIELSEKTVIQTWHKATQKQDPVAILKRLQKEKLLAT----- 914
 QY 76 QGGPGEDGAGQEPAPEDAKSKRTYVAR-NGE--PEPTPVYVGEKEPSK-----GDPNTESE 128
 DB 915 -----EQSDAAVAKSKRLNELNEMASEKAKAAAGEAKVKQQLVARQEI 960
 QY 129 IQSDVWGRDHRRPQEKKKAK-----GLGKETITLLMQLN----- 164
 DB 961 VQARQASVYRDVHVKVQQLQKIRTLQEQLENGPNTQLARQENSILRDALNATQSOVE 1020
 QY 165 -----TLSPTEKLAALCKKYAELLEHNRSOKWKLKQKQSQVQSKDHLGERSKAV 219
 DB 1021 SKQNTLAKLRLQELSKVNLKELVEKSEASQEQKQKALEAKAATFEKVLQQAQSHKSE 1080
 QY 220 LA-RSKLESCLREL---ORHNESLKEEGVQVAREEKKKEVTSHTFQVTLNDIQLQMEQH 275
 DB 1081 EALQRLKEVTELCRAQTSNANRAD-AKRAQEQQORVAELHSLKQSSSEVSKSCEEL 1139
 QY 276 NERNESKL---ROENMELAEFLKLIQYELREBEHDKVFKHKLQQLQVDAK----- 324
 DB 1140 SSLHQLKEARNSQLTIRISIALLEAGQAQ-DTQASHAEANQQQTLKELESQVSC 1198
 QY 325 LQAQEMLKEAERHQ-----REKDFLLKAVE-SQRCMLMKQOETHLKKQ----- 370
 DB 1199 LEKFTSELKEVEQKQKNDRLRKNKAWALAEACERACEKRLSLTQAKSESEKQLHL 1258
 QY 371 -----LAL-----YTEKPEEF----- 381
 DB 1259 AEAQTKETLALLPGLSISAHQNYAEWLQEFKEKSELKKPPTLEPSMDIVLKLREAE 1318
 QY 382 -QNTLSKSSEVTTFKQEMKQTKIKK-LEKETTMYRSRWSSNKALIEAEKTVRK 439
 DB 1319 TQNSLQAECDOYRTILAETEGMLKDLQKSVSEERV-----WKAKVGAABEELHKSRTVK 1374
 QY 440 ELEGLOVKIQ-----RLEKLCRALQTERNDLNKRV---QDLASAGQG 478
 DB 1375 HLEDIVELKLGHESSDQVREHTSHLEAELEKHMMAASAECQNYAKEVAGLRQLLESQS 1434
 QY 479 SUTDSQPEPR-----PQPGQAQAPSSPRVTEAP 506
 DB 1435 QLDEAKSEAKQSGDELALVRQQLSDMRSHVEDGVAGSPVAPPAQDP 1482
 RESULT 38
 ID MYH6_MESAU STANDARD; PRT; 1939 AA.
 AC F13539; Q60562;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin heavy chain, cardiac muscle alpha isoform (MyHC-alpha).
 GN MYH6.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;

RN SEQUENCE FROM N.A.
 RC STRAIN=FIB; TISSUE=Liver;
 RX MEDLINE=95115033; PubMed=7815459;
 RA Wang R., Sole M.J., Cukerman E., Liew C.-C.;
 RT "Characterization and nucleotide sequence of the cardiac alpha-myosin
 heavy chain gene from Syrian hamster";
 RL J. Mol. Cell. Cardiol. 26:1155-1165(1994).
 [2]
 RN SEQUENCE OF 1630-1939 FROM N.A.
 RX MEDLINE=86205859; PubMed=3458174;
 RA Liew C.-C., Jandreski M.A.;
 RT "Construction and characterization of the alpha form of a cardiac
 myosin heavy chain cDNA clone and its developmental expression in the
 Syrian hamster";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:3175-3179(1986).
 CC
 CC !- FUNCTION: Muscle contraction.
 CC !- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
 heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
 and 2 regulatory light chain subunits (MLC-2).
 CC !- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC !- DOMAIN: The rodlike tail sequence is highly repetitive, showing
 cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
 characteristic for alpha-helical coiled coils.
 CC !- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
 meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be
 split further into 2 globular subfragments (S1) and 1 rod-shaped
 subfragment (S2).
 CC !- MISCELLANEOUS: The cardiac alpha isoform is a 'fast' ATPase
 myosin, while the beta isoform is a 'slow' ATPase.
 CC !- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC !- SIMILARITY: Contains 1 IQ domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L15351; AAB59701.1; -;
 DR EMBL; M12995; AAA37081.1; -;
 DR PIR; I48175; I48175.
 DR HSP; P08799; I48175.
 DR InterPro; IPR000048; IQ region.
 DR InterPro; IPR001609; myosin head.
 DR InterPro; IPR004009; Myosin N.
 DR InterPro; IPR002928; Myosin tail.
 DR Pfam; PF00612; IQ; 2.
 DR Pfam; PF00963; myosin head; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS00096; IQ; 1.
 DR
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 FT ATP-binding; Methylation; Multigene family; Calmodulin-binding.
 FT DOMAIN 783 812
 FT MYOSIN HEAD-LIKE.
 FT COILED COIL (POTENTIAL).
 FT POLY-ALA.
 FT ATP (POTENTIAL).
 FT ACTIN-BINDING.
 FT METHYLATION (TRI-) (POTENTIAL).
 FT MOD RES 129 129
 FT Q -> L (IN REF. 2).
 FT CONFLICT 1633 1633
 FT H -> Q (IN REF. 2).
 FT EL -> DV (IN REF. 2).
 FT V -> G (IN REF. 2).
 FT CONFLICT 1693 1693
 FT K -> R (IN REF. 2).
 FT CONFLICT 1844 1844

FT CONFLICT 1879 1879 A -> T (IN REF. 2).
 FT CONFLICT 1885 1885 E -> Q (IN REF. 2).
 FT CONFLICT 1907 1907 E -> V (IN REF. 2).
 FT CONFLICT 1928 1928 D -> N (IN REF. 2).
 FT CONFLICT 1933 1933 QKM -> KR (IN REF. 2).
 SQ SEQUENCE 1939 AA; 223626 MM; DBC8297DFE83115A CRC64;
 Query Match 8.5%; Score 230.5; DB 1; Length 1939;
 Best Local Similarity 24.4%; Pred. No. 0.027;
 Matches 146; Conservative 80; Mismatches 197; Indels 175; Gaps 25;
 QY 12 EGAQRPSCAPAVAEAGPGSSQAPRKQEGQAQTAQSGALRDVSEELSRQLEDILSTYC 71
 DB 1037 EGSLEQKVKRMVLE-----RAKKLEGLDNLVTOSS--INDLEND-XLQLEEKKKKE 1086
 QY 72 VDNNOGPGGEDCAQGEPAEPEDA-----EKSRTYVAR-----NCEPEPTPVVYGEKEP 119
 DB 1087 FDISQ-----QNSKLEDEQALALQKQKLEKQARIEELEEELEAERTAKVKEKL 1138
 QY 120 SKGDPNTHBIROS-DEVG-----DRHRRPQEKKAKGLCKETLLMQTLNTLTSTPEE 171
 DB 1139 SDLTRELEIEISERLEEAGGATSVQIEMNKYKRAEFQKMRDUEATL-----QHEA 1189
 QY 172 KLAALCKKY---AELEERHSQKMKLLQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 227
 DB 1190 TAALEKKAHDSVAELGEQIDNLQKVKQKLEKSEKFLDLDVTSNMQIHKAKANLEK 1249
 QY 228 LCR-----ELQRNRSLSKEGVQVAP-----ELQRNRSLSKEGVQVAP----- 248
 DB 1250 VSTLEDOQAEYRVKLEESQSLNDPTQKALQKQKQKQKQKQKQKQKQKQKQKQKQK 1309
 QY 249 -----EEBKREKVTSH-FOVTLNDIQIQMEQHNER-----NSK 281
 DB 1310 YTOQMEDLKRQLEEEGAKKALAHALQSAHDCDLAREQYEEBEWEAKAELQKRVLSKANSE 1369
 QY 282 LR-----QENNELAELRLKLIQEVLEBEEHIDKY-----FKHK---DLQK 318
 DB 1370 VAQWRTKYETDAIQRTLELEEAQKQKLAQRIQDAEEAVEAVNAKCSLEKTKHQLONEIED 1429
 QY 319 QLYDAKLOQAQEMLKABSEHGHQEKDFLLKEAVESQRMCELMKQOETHKQQLALATYEFK 378
 DB 1430 LAMVDERSNA---AAALQK 1484
 QY 379 EEFQNTLSKSSRYFTTFKQEMEQTKI-----KKLEKETTMYR 417
 DB 1485 -KLKNAYEESLEHETFKRENKQLQEEISDLTEQLGEGGKNNVHELEKRVKQLEVKQKQKQK 1543
 QY 418 SRWESSNKALLEMAKRYVDKKEGLQVKIQKLEKLC-----RALQTERNDLNKRVOD 471
 DB 1544 SALEEA-EASLEHEBGLIR-AQLSEFQIKASIERKLAKEDEEMKQAKRNHL--RVVD 1597
 RESULT 39
 ID AKAP9_HUMAN STANDARD; PRT; 3911 AA.
 AC Q95996; O14869; O43355; Q94895; Q9UOH3; Q9UQ04; Q9Y6B8; Q9Y6Y2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 15-MAR-2004 (Rel. 43, Last annotation update)
 DE A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)
 DE (PKA9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor
 protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)
 DE (Hyperion protein) (Yotiao protein) (Centrosome- and Golgi-localized
 PKM-associated protein) (CG-NAP).
 DE AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.
 GN Homo sapiens (Human).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A. (ISOFORM 4).
 RP TISSUE=Brain;
 RC MEDLINE=98151389; PubMed=9482789;
 RX

RA Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.,
RT "Yotiao, a novel protein of neuromuscular junction and brain that
RT interacts with specific splice variants of NMDA receptor subunit
RT NR1.",
RL J. Neurosci. 18:2017-2027(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.
RX MEDLINE=99219864; PubMed=10202149;
RA Witczak O., Skaalhegg B.S., Keryer G., Bornens M., Taeken K.,
RA Jahnson T., Oerstavik S.,
RT "Cloning and characterization of a cDNA encoding an A-kinase anchoring
RT protein located in the centrosome, AKAP450.",
RL EMBO J. 18:1858-1868(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Brain;
RX MEDLINE=99287934; PubMed=10358086;
RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.,
RT "Characterization of a novel giant scaffolding protein, CG-NAP, that
RT anchors multiple signaling enzymes to centrosome and the Golgi
RT apparatus.",
RL J. Biol. Chem. 274:17267-17274(1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Kemmer W.A., Deiss S., Schwarz U.,
RT "Cloning of Hyperion.",
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).
RC TISSUE=Gastric parietal cell;
RX MEDLINE=99115654; PubMed=9915845;
RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,
RA Trotter K.W., Milgram S.L., Goldenring J.R.,
RT "AKAP350, a multiply spliced protein kinase A-anchoring protein
RT associated with centrosomes.",
RL J. Biol. Chem. 274:3055-3066(1999).
RN [6]
RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
RC TISSUE=Lymphoblast;
RX MEDLINE=99115654; PubMed=9915845;
RA Hinds K., Sutterer C., Becker M., Hawkins M.,
RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).
RC TISSUE=Lung;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.,
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.",
RL DNA Res. 5:277-286(1998).
RN [9]
RP SEQUENCE OF 17-1800 FROM N.A.
RA Wu X., Graves T., Bradshaw H.,
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Binds to type II regulatory subunits of protein kinase
CC A. Scaffolding protein that assembles several protein kinases and
CC phosphatases on centrosome and Golgi apparatus where physiological
CC events can be regulated by phosphorylation state of protein
CC substrate. Isoform 4/Yotiao is associated with the N-methyl-D-
CC aspartate receptor and is specifically found in the neuromuscular
CC junction (NMJ) as well as in neuronal synapses explaining that its
CC role may be to organize postsynaptic specializations.
CC -1- SUBUNIT: Interacts with the regulatory region of protein kinase N
CC (PKN). Interacts with protein phosphatase 1 (PP1)
CC and the immature non-phosphorylated form of PKC epsilon.
CC

CC -1- SUBCELLULAR LOCATION: Centrosomal in many cell types and
CC cytoplasmic in parietal cells.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=6;
CC Name=1;
CC IsoId=Q99996-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q99996-2; Sequence=VSP_004102, VSP_004107;
CC Name=3; Synonyms=CG-NAP;
CC IsoId=Q99996-3; Sequence=VSP_004102, VSP_004105, VSP_004107;
CC Name=4; Synonyms=Yotiao;
CC IsoId=Q99996-4; Sequence=VSP_004103, VSP_004104;
CC Name=5;
CC IsoId=Q99996-5; Sequence=VSP_004108;
CC Name=6; Synonyms=AKAP350;
CC IsoId=Q99996-6; Sequence=VSP_004106, VSP_004107, VSP_004109;
CC -1- TISSUE SPECIFICITY: Widely expressed. Isoform 4/Yotiao is highly
CC expressed in skeletal muscle and in pancreas.
CC -1- DOMAIN: RI-binding site, predicted to form an amphipathic helix,
CC could participate in protein-protein interactions with a
CC complementary surface on the R-subunit dimer.
CC -1- CAUTION: Ref.6 sequence differs from that shown due to two
CC frameshifts in positions 3782 and 3811.
CC -1- CAUTION: Ref.9 sequence differs from that shown due to four
CC frameshifts in positions 29, 1653, 1699 and 1735.
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

EMBL; AJ131693; CAB40713.1; --	DR	EMBL; AJ131693; CAB40713.1; --	DR
EMBL; AB019691; BAA78718.1; --	DR	EMBL; AB019691; BAA78718.1; --	DR
EMBL; AJ010770; CAA0361.1; --	DR	EMBL; AJ010770; CAA0361.1; --	DR
EMBL; AF026245; AAB86384.1; --	DR	EMBL; AF026245; AAB86384.1; --	DR
EMBL; AF03037; AAD22767.1; --	DR	EMBL; AF03037; AAD22767.1; --	DR
EMBL; AC004013; AAB96867.1; ALT_FRAME	DR	EMBL; AC004013; AAB96867.1; ALT_FRAME	DR
EMBL; AF091711; AAD37919.1; --	DR	EMBL; AF091711; AAD37919.1; --	DR
EMBL; AB018346; BAA34523.1; --	DR	EMBL; AB018346; BAA34523.1; --	DR
EMBL; AC000066; AAC60380.1; ALT_FRAME	DR	EMBL; AC000066; AAC60380.1; ALT_FRAME	DR
Genew; HGNC:379; AKAP9.	DR	Genew; HGNC:379; AKAP9.	DR
MTM; 604001; --	DR	MTM; 604001; --	DR
GO; GO:0005813; C:centrosome; TAS.	DR	GO; GO:0005813; C:centrosome; TAS.	DR
GO; GO:0005856; C:cytoskeleton; TAS.	DR	GO; GO:0005856; C:cytoskeleton; TAS.	DR
GO; GO:0004973; F:N-methyl-D-aspartate receptor-associated pr. . . ; TAS.	DR	GO; GO:0004973; F:N-methyl-D-aspartate receptor-associated pr. . . ; TAS.	DR
GO; GO:0005515; F:protein binding; TAS.	DR	GO; GO:0005515; F:protein binding; TAS.	DR
GO; GO:0007165; P:signal transduction; TAS.	DR	GO; GO:0007165; P:signal transduction; TAS.	DR
GO; GO:0007268; P:synaptic transmission; TAS.	DR	GO; GO:0007268; P:synaptic transmission; TAS.	DR
GO; GO:0006810; P:transport; TAS.	DR	GO; GO:0006810; P:transport; TAS.	DR
Coiled coil; Alternative splicing; Polymorphism.	DR	Coiled coil; Alternative splicing; Polymorphism.	DR
DOMAIN 2554 2567 PRA-RII SUBUNIT BINDING DOMAIN.	KW	DOMAIN 2554 2567 PRA-RII SUBUNIT BINDING DOMAIN.	KW
DOMAIN 164 934 COILED COIL (POTENTIAL).	FT	DOMAIN 164 934 COILED COIL (POTENTIAL).	FT
DOMAIN 944 1022 COILED COIL (POTENTIAL).	FT	DOMAIN 944 1022 COILED COIL (POTENTIAL).	FT
DOMAIN 1100 1185 COILED COIL (POTENTIAL).	FT	DOMAIN 1100 1185 COILED COIL (POTENTIAL).	FT
DOMAIN 1253 1280 COILED COIL (POTENTIAL).	FT	DOMAIN 1253 1280 COILED COIL (POTENTIAL).	FT
DOMAIN 1336 1392 COILED COIL (POTENTIAL).	FT	DOMAIN 1336 1392 COILED COIL (POTENTIAL).	FT
DOMAIN 1434 1459 COILED COIL (POTENTIAL).	FT	DOMAIN 1434 1459 COILED COIL (POTENTIAL).	FT
DOMAIN 1585 1659 COILED COIL (POTENTIAL).	FT	DOMAIN 1585 1659 COILED COIL (POTENTIAL).	FT
DOMAIN 1857 2455 COILED COIL (POTENTIAL).	FT	DOMAIN 1857 2455 COILED COIL (POTENTIAL).	FT
DOMAIN 2544 2561 COILED COIL (POTENTIAL).	FT	DOMAIN 2544 2561 COILED COIL (POTENTIAL).	FT
DOMAIN 2603 2776 COILED COIL (POTENTIAL).	FT	DOMAIN 2603 2776 COILED COIL (POTENTIAL).	FT
DOMAIN 3065 3092 COILED COIL (POTENTIAL).	FT	DOMAIN 3065 3092 COILED COIL (POTENTIAL).	FT
DOMAIN 3124 3470 COILED COIL (POTENTIAL).	FT	DOMAIN 3124 3470 COILED COIL (POTENTIAL).	FT
DOMAIN 3587 3689 COILED COIL (POTENTIAL).	FT	DOMAIN 3587 3689 COILED COIL (POTENTIAL).	FT
DOMAIN 3726 3730 POLY-LEU.	FT	DOMAIN 3726 3730 POLY-LEU.	FT
DOMAIN 203 292 GLN-RICH.	FT	DOMAIN 203 292 GLN-RICH.	FT
DOMAIN 321 1010 GLU-RICH.	FT	DOMAIN 321 1010 GLU-RICH.	FT
DOMAIN 1846 2772 Missing (in isoform 2 and isoform 3).	FT	DOMAIN 1846 2772 Missing (in isoform 2 and isoform 3).	FT
VARSPLIC 17 28 /FTId=VSP_004102.	FT	VARSPLIC 17 28 /FTId=VSP_004102.	FT

Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D., Kani A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maitais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sanderlin A., Schneider C., Semple C.A., Setou M., Shinada K., Sultana R., Tanaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh N., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.,

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs." ;

-!- FUNCTION: Binds phospholipid vesicles containing phosphatidylinositol 3-phosphate and participates in endosomal trafficking (By similarity).

-!- SUBUNIT: Homodimer. Binds STX6. Binds RAB5A, RAB5B, and RAB22A that have been activated by GTP-binding (By similarity).

-!- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein associated with early endosomes (By similarity).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement.shtml>).

Query Match	8.5%;	Score	229.5;	DB	1;	Length	790;
Best Local Similarity	22.1%;	Pred. NO.	0.012;				
Matches	110;	Conservative	92;	Mismatches	196;	Indels	99;
Gaps	17;						
QY	52	LRDVSE----	ELSRQLEDILSTY	CVDDNNOG	PGDGGAGCEP	APEDPAEKSR	TVVANGEP
DB	1	MKDLPEKQAQLATE	IADIKSY-----	-----	DEEKS	LRAAAAEKV	THLTEDLANK
QY	108	EPTPVVYGEKEPS	GDPNTEIR-----	-----	QSDEVGDR	DRHRPQSEK	KAAGLGE
DB	47	Q-TTVTQDLTELL	QRPGEDVAIL	KKELVQVOTL	MDNMTLE	RERSEK	LDCCKKLOSE
QY	156	ITLLMOTLNTLS	PEEKLAALCK	YALFEEHRS	QOMKLLQK	QSOLVQEKD	----
DB	106	HAHLEATING	RLSELAKPQ	EVAVYQEI	QKGLSINELT	QKNQLTEK	QKQKDLDTYHL
QY	212	RGEHSAVLA	RSKLSEI-----	-----	CHRELQ-----	-----	-----
DB	166	EERKHNSASRKT	QLASLHORD	LDCCQLQARL	PASSESSLO	RAGEUSEK	EAARAKLEEL
QY	252	EKKKEVTSHPQ	VTINDIQLQ----	-----	MEOR-----	-----	-----
DB	226	REVSTFRQLK	VEVKQLQQ	QREKQHG	LQGEVSL	CHKLLFET	ERQLGEAHGELKEQR
QY	301	ELREHIDKVF	KHDKLQQQ	LVD-----	AKLQQAQEM	LKE-----	-----
DB							
QY							

```

DB      1536 QLDHXSQTSLEAEAA 1553
|||||   |||||   |||
RESULT 41
EEAL MOUSE
ID EEAL MOUSE STANDARD; PRG; 790 AA.
AC QBBL66;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Early endosome antigen 1 (Fragment).
GN EEAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
EN
FP SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
FA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaoka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldaelli R., Hill D.P., Bult C., Hume D.A., Schackenbush J.,
RA Schrim L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
FA Slater J.A., Bradt D., Brusic V., Choitha C., Corbani L.E., Cousins S.,

```



```

QY 286 -NWEAERUKKLEIYQVELREBHDKV-----FKYDLOOOL-----VDKLLQQAQ 329
DB 664 FOAHIEEKNVETKLEKLDVQTELESLSSELSVLKARHK-LEELSVLKQDQDKMKQBLE 722
QY 330 EMLKAEERHOREKOPFLKAEVSRMCB-----LMKQOETHKQ----- 369
DB 723 AKMDQKQHEHQOVDSIIKEHVSIOETKALKDQINQLELLKDKRDLKHEHQAVENL 782
QY 370 -----OLALYTKFEPFQNTLSKSEVFTTFKQEMKTKIKKLEKTTWYRSRW 420
DB 783 EADIKSEGELOQASAKLDFVOSYATHEQTAYEEQLAQOQLDLETERILL----- 838
QY 421 ESSNKALLWEAKETVRDKELEGLOVKIORLEKLCALOTERNDLKRVQDLSAGQGSIL 480
DB 839 ----TKQVAEVAQKEVCTELDAHKIQVQ---DLAQOLEKQSEKQVKSLTVYBSKL 892
QY 481 TDGSPER 487
DB 893 EDGNKEQ 899

RESULT 43
GOM4_MOUSE
ID GOA4_MOUSE STANDARD; PRT; 2238 AA.
AC Q91VW5; O70365; Q8CGH6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Golgi autoantigen, golgin subfamily A member 4 (tgolgin-1).
GN GOLGA4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=22153218; PubMed=12162805;
RA Cowan D.A., Gay D., Bieler B.M., Zhao H., Yoshino A., Davis J.G.,
RA Tomayko M.M., Murali R., Greene M.I., Marks M.S.;
RA "Characterization of mouse Tgolgin-1 (golgin-245/trans-golgi p230/256
RT kd golgin) and its upregulation during oligodendrocyte development.";
RL DNA Cell Biol. 21:505-517(2002).
RN [2]
SEQUENCE OF 1-980 AND 991-2231 FROM N.A., AND VARIANTS ARG-61;
RP SER-280; GLU-293; SER-638; THR-819; THR-829 AND ALA-859.
RC STRAIN=C57BL/6, and Czech II; TISSUE=Brain, and Breast cancer;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bonak S.A., McWay P.J., McKernan K.J., Malek J.A., Guarnatone P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting R.W., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Sutterch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: May play a role in vesicular transport from the trans-
Golgi (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic, peripheral membrane protein
associated with the trans-Golgi network.
CC -!- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in

```

```

CC oligodendrocyte precursors, particularly at a stage just prior to
myelination.
CC -!- DOMAIN: Extended rod-like protein with coiled-coil domains.
CC -!- SIMILARITY: Belongs to the golgin family.
CC -!- SIMILARITY: Contains 1 GRIP domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF051357; AAC05573.2; --
DR EMBL; BC007485; AAH07485.1; --
DR EMBL; BC037641; AAH37641.1; ALT_INIT.
DR EMBL; BC053000; AAH53000.1; --
DR PIR; T14265; T14265.
DR MGD; MGI:1859646; Golga4.
DR GO; GO:0005794; C:Golgi apparatus; IDA.
DR InterPro; IPR000237; GRIP domain.
DR InterPro; IPR003345; M_repeat.
DR Pfam; PF01465; GRIP; 1.
DR Pfam; PF02370; M; 12.
DR PROSITE; PS0913; GRIP; 1.
KW Golgi stack; Coiled coil; Polymorphism.
FT DOMAIN 156 2161 COILED COIL (POTENTIAL).
FT DOMAIN 335 1011 GLU-RICH.
FT DOMAIN 1855 2107 GLU-RICH.
FT DOMAIN 2178 2225 GRIP.
FT VARIANT 61 61 T -> R (in strain Czech II).
FT VARIANT 280 280 G -> S (in strain Czech II).
FT VARIANT 293 293 G -> E (in strain Czech II).
FT VARIANT 638 638 G -> S (in strain Czech II).
FT VARIANT 819 819 K -> T (in strain Czech II).
FT VARIANT 829 829 A -> T (in strain Czech II).
FT VARIANT 859 859 T -> A (in strain Czech II).
FT VARIANT 2065 2065 R -> S (in strain C57BL/6).
FT SEQUENCE 2238 AA; 257562 MW; 494EA2C11F0165B CRC64;
QY Query Match 8.5%; Score 229.5; DB 1; Length 2238;
Best Local Similarity 22.1%; Pred. No. 0.034;
Matches 119; Conservative 107; Mismatches 200; Indels 113; Gaps 19;
QY 15 QERPQAPAVEAEFGSGSQAPRKPEGAQAR-----TAQSGALRDVSE-----ELSR 61
DB 424 QKEKSERAAFELEKALST--AQKTEDAQRMKVMDEQMKAVESBEERLRLQHELSR 481
QY 62 QLEDILSTVCYDNNQSGFGDGAQGEPAEPEDAEKSRITTVARNQSPETPVVYGEKEPSK 121
DB 482 VQEAASVAKKNSEQVAAQLKHAELASKEQLSRLEAREERELQEQMRITALEKSRSE 541
QY 122 GDPNTEIRQSDVSG-----DRHRRPQKKKAKGLGKEI-----TLLMOTLN 164
DB 542 YLKLTOEKEQESLALSELELQKKAILESENKLOELQGEAEAYRTRILEFTLEKSLQ 601
QY 165 TLSTPEELALC-----KKYAELEERHNSQKMKLLQKQSQVQEKDHLRGHS 216
DB 602 ESKTQSEHLAVHLEAKKKNKELTALAEQHT---EVEGLQQQDQLWTE----- 649
QY 217 KAVLARSKLESICRELQHRNLSKEGVQCPARFEEBKVKVTSHPQVTLNDI-QLOMEQH 275
DB 650 -----ELQSL-----QQHQAAVEELRSKYQOEKALLKEKESLFQAHIQDNVETKEL 699
QY 276 NERNSKLQENNELAERLK---KLTEQVELREEDTKVFKHKDLOQLVADKLQQAQOE-- 330
DB 700 DKQKQELSVSSELSEALFARDQLABELSVLRGDADKM---KQALEAELEQRHH 752
QY 331 -----MLKAEERHOREKDFLKAEVSRMCBQKQETHLK-----QQALYTKFPE 379
DB 753 QREVSGISQOQLTVRAEKALKD--ELSLGALLDERDEHLRERQARQVDLEAHLQKSA 810

```

QY 380 -FQNTLSK-----SSEVFTTFKQEMKRYKKIKLEKTTMYRSRVESNKALL 428
Db 811 GELQALAKLILLHSQSAAREQAGAYEBQLAQMOOKVLDLETEKSL-----TKQV 863
QY 429 ENABEKTVDEKELEGVQIKRLEKLCRALQTERNDLNKRVQDLSAGGQGSLTDSGPER 487
Db 864 EMETHKGHVCELDQAQAVQQLER-----QRSELEKVRSLAQQLDSQLKNSTVEK 915

RESULT 44
MYHA HUMAN
ID MYHA HUMAN STANDARD; PRT; 1976 AA.
AC P35580;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain, type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).
GN MYH10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96025307; PubMed=7499478;
RA Phillips C.L., Yamakawa K., Adelstein R.S.;
RT "Cloning of the cDNA encoding human nonmuscle myosin heavy chain-B and analysis of human tissues with isoform-specific antibodies.";
RL J. Muscle Res. Cell Motil. 16:379-389(1995).
RN [2]
RP SEQUENCE OF 63-722 FROM N.A.
RX MEDLINE=91316803; PubMed=1860190;
RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K., Gdula D., Adelstein R.S., Weir L.;
RT "Human nonmuscle myosin heavy chains are encoded by two genes located on different chromosomes.";
RL Circ. Res. 69:530-539(1991).
CC -!- FUNCTION: Cellular myosin appears to play a role in cytokinesis, cell shape, and specialized functions such as secretion and capping.
CC -!- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2 regulatory light chain subunits (MLC-2).
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing cycles of a 28-residue repeat pattern composed of 4 heptapeptides, characteristic for alpha-helical coiled coils.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
CC
CC EMBL; M69181; AAA99177.1; -;
CC PIR; A59252; A59252.
CC HSSP; P10587; 1BR2.
CC Genes; HGNC:7568; MYH10.
CC MIM; 160776; -;
CC GO; GO:0016459; C:myosin; NAS.
CC GO; GO:0003779; F:actin binding; NAS.
CC GO; GO:0005524; F:ATP binding; NAS.
CC GO; GO:0000910; P:cytokinesis; NAS.
CC InterPro; IPR000048; IQ region.
CC InterPro; IPR001609; myosin head.
CC InterPro; IPR004009; Myosin_N.
CC InterPro; IPR002928; Myosin_tail.
CC InterPro; IPR002017; Spectrin.
CC Pfam; PF00612; IQ; 1.

Pfam; PF00063; myosin head; 1.
Pfam; PF02736; Myosin_N; 1.
Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PF00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin, ATP-binding; Calmodulin-binding; Actin-binding; Coiled coil;
FT DOMAIN 1 785 MYOSIN HEAD-LIKE.
FT DOMAIN 786 815 IQ.
FT DOMAIN 845 1976 COILED COIL (POTENTIAL).
FT NP BIND 178 185 ATP (POTENTIAL).
SQ SEQUENCE 1976 AA; 228938 MW; B2BB87FF35EA124F CRC64;
Query Match 8.5%; Score 229; DB 1; Length 1976;
Best Local Similarity 24.2%; Pred No. 0.031;
Matches 93; Conservative 76; Mismatches 116; Indels 100; Gaps 16;
QY 160 MDTLNTLTPEEKLAALCKYAELEHNRNSQKMKLQKXOSQLVQSKDH----- 211
Db 841 VKPLQVTRQEEELQAKDELLKVKQKTKVEGELEEMERKHQQLLEKNILAEQLQAE 900
QY 212 -----RGEHSKAVLARSK--LESCLRELQHRNSLKEEGVQARAREEK-----RKEVT 258
Db 901 ELFAEAEEMRRLAALAKQELSEILHDL-----SRVEEERNRQILQNEKKWQ 949
QY 259 SHFQVTLNDIQLOMEQHNRNSKLEQNMELAEKLIQVYELREHHDKVKFKHQLQQ 318
Db 950 AHIQ-----DLSEQLDEBEGARQLQLEKTAETAKIKKWESEILLLEDQNSKFKKELME 1005
QY 319 -----QLVD-----AKLQQAQ--MLKEAEERHORE-----KDFLLKSAV 351
Db 1006 DRIARCSQLAEERKAKNLAIRKRYKVMISDLERLKEEKTRQELEKAKRKLQGETT 1065
QY 352 ESQRMCELMKQOETHLKOQALALYTEKFBFQNTLSKSEVFTTFKQEMKMTK----KIT 407
Db 1066 DLQDQIAELQAIQIDELQLQLA---KKESELOQALARGDDE-TLKNNAKLVVRELQAQIA 1121
QY 408 KLEKETTMYRSRVESNKALLEAEKTVRD--KELEGLOVKI----- 448
Db 1122 ELQED---PESEKASRNKA-----EKQRDLSELEALKTELEDTLDTAAQQLERTKR 1172
QY 449 -QLEKLCALQTERNDLNKRVQDL 472
Db 1173 EQVAELKKALEETKYNHEAQIQDM 1197

RESULT 45
LVA_DROME
ID LVA_DROME STANDARD; PRT; 2779 AA.
AC Q8MS51; Q8W4N7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Lava lamp protein.
GN LVA OR CG6450.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F., George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.X., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.A., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dedson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalueh F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lei Z.,
RA Lasco P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy E., Murphy L., Muzzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K.A., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimani I., Simpson M., Skupski K.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE OF 980-2779 FROM N.A.
RC STRAIN=Berkeley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarini H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celinker S.E.;
RT "A *Drosophila* full-length cDNA resource."
RL Genome Biol. 3:RESEARCH080.1-RESEARCH080.8(2002).
RN [3]
RP FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=20530668; PubMed=11076973;
RA Sisson J.C., Field C., Ventura R., Royou A., Sullivan W.;
RT "Lava lamp, a novel peripheral golgi protein, is required for
RT *Drosophila* melanogaster cellularization."
RL J. Cell Biol. 151:905-918(2000).
CC -!- FUNCTION: Lva and spectrin may form a Golgi-based scaffold that
CC mediates interaction of Golgi bodies with microtubules and
CC facilitates Golgi-derived membrane secretion required for the
CC formation of furrows during cellularization.
CC -!- SUBUNIT: Interacts with Clp-190 and spectrin separately.
CC -!- SUBCELLULAR LOCATION: Lva-alpha-spectrin and Lva-CLIP-190
CC complexes are found at the Golgi.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AR003429; AAF45910.1; -
CC EMBL; AY118638; AAM50007.1; ALT_INIT.
CC FlyBase; FBgn0029688; lva.
CC
CC Developmental protein; Golgi stacked; Coiled coil.
CC
CC DOMAIN 52 85 COILED COIL (POTENTIAL).
CC FT DOMAIN 141 175 COILED COIL (POTENTIAL).
CC FT DOMAIN 220 507 COILED COIL (POTENTIAL).
CC FT DOMAIN 559 716 COILED COIL (POTENTIAL).
CC FT DOMAIN 751 1733 COILED COIL (POTENTIAL).
CC FT

FT DOMAIN 1785 1863 COILED COIL (POTENTIAL).
FT DOMAIN 1941 2433 COILED COIL (POTENTIAL).
FT DOMAIN 2504 2544 COILED COIL (POTENTIAL).
FT DOMAIN 2600 2641 COILED COIL (POTENTIAL).
FT CONFLICT 1211 1211 E -> K (IN REF. 2).
FT CONFLICT 1424 1424 A -> S (IN REF. 2).
FT CONFLICT 1506 1506 E -> D (IN REF. 2).
FT CONFLICT 1534 1534 A -> V (IN REF. 2).
FT CONFLICT 2034 2034 G -> H (IN REF. 2).
FT CONFLICT 2059 2059 Q -> E (IN REF. 2).
FT CONFLICT 2133 2133 Q -> E (IN REF. 2).
FT CONFLICT 2154 2154 E -> V (IN REF. 2).
FT CONFLICT 2160 2160 L -> P (IN REF. 2).
FT CONFLICT 2200 2200 A -> T (IN REF. 2).
FT CONFLICT 2217 2217 E -> D (IN REF. 2).
FT CONFLICT 2271 2271
SQ SEQUENCE 2779 AA; 315897 MW; 1CB3965102018AEE CRC64;
Query Match 8.5%; Score 229; DB 1; Length 2779;
Best Local Similarity 21.7%; Pred. No. 0.043;
Matches 127; Conservative 98; Mismatches 223; Indels 136; Gaps 20;
QY 5 GQPEAGPEGAGQERPSQAP-----AVRAGPGSSQAPKPEGAQARTASGA 51
DB 619 QRRPASBGAQEQKQVHPQPSHVSELTQTEDEDSGETLSQURRELELFTQERGE 678
QY 52 LRDVSEELS-----ROLF-DILSTYCYDNNQGGFGEDGAQGEPA 89
DB 679 VLDKLEQLSAENLQALRESSSLQLOKREKOLISTSTSSNLS-----Q 727
QY 90 EPEDAESRTVAARNGPPTPVYVG--EKSPKGDPTBEIRCSDEVGDRDHRPQKK 147
DB 728 ELSNQRSSSEVVATLDAGEGPNLFKCEKSLKSLSELAAYRANRQAKFNVSXKLAK 787
QY 148 KAKGLGKEITLIMQTLNLTSTP-----EKKALCKYAELEFHNRSOKMKLLOKKQ 201
DB 788 EAKNCHTQLSELLHKVKEASTAVETVTVVWVATPANGKALAEYQLNAQNAELKAVI 847
QY 202 SOLVQEKDHLRGEH--SKANLA-----RSKLSLCRELORHNSLKEEYV 244
DB 848 SKLQELDELRESYPTEAPLAIWGSQSQRDEILQLOQLED-ARSLQAEQKQIEQV 906
QY 245 ---QRAREEERKKEVTSHFQVTLNDIQLQMEQHNE-RNSKLQENNELAKLKLIEQY 300
DB 907 DOIKELRQTEAEQQLVARQSAEITQLQSEQPDQLNSKMSHEKQL-EQOTRIRREL 965
QY 301 ELREHIDKVPK-----HKLOQQLQVDA-----KLQQAQEMLKE---AEERH 339
DB 966 EARAESLEGELSLIQLTVAEQKQOLIEVSSESEHALNKLMLQSAQBELRELRAKEDPD 1025
QY 340 QREKDFLLKEAVESQRMCKMQOET--HLKQQLALY-----TEKFEFQV--- 383
DB 1026 QLRALRVSKSLVAQVQVELTSSQETVDALNQIQEYQGLEHAHKEBQFNRLREKLKK 1085
QY 384 -----TLSSKSEVFTTFKQEMKMTKIKLKEKTTMYRSRWSSESSKALLEM 430
DB 1086 YALNKKRTQDNADLEQKQVQLTSQLOEQBELVKQKEEVERPIVDNHRVQLQQQVSKL 1145
QY 431 AEEKTV-----RDKELEGVLVKIQLREKLCLALQTERNDLN 466
DB 1146 NEDLKAKIHLNENR-ALRQLKQIQEQQLQERDAELQDAN 1188
RESULT 46
MYS_AEQIR
ID MYS_AEQIR STANDARD; PRT; 1938 AA.
AC P24733;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, striated muscle.
OS Aequipecten irradians (Bay scallop).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pectinoida;
OC Pectinoidea; Pectinidae; Argopecten.

OX NCBI_TaxID=31199;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adductor muscle;
RX MEDLINE=92011595; PubMed=1917970;
RA Nyitray L., Goodwin E.B., Szent-Gyorgyi A.G.;
RT "Complete primary structure of a scallop striated muscle myosin heavy
chain. Sequence comparison with other heavy chains reveals regions
that might be critical for regulation.";
RL J. Biol. Chem. 266:18469-18476(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Adductor muscle;
RX MEDLINE=9108319; PubMed=2263488;
RA Nyitray L., Goodwin E.B., Szent-Gyorgyi A.G.;
RT "Nucleotide sequence of full length cDNA for a scallop striated
muscle myosin heavy chain.";
RL Nucleic Acids Res. 18:7158-7158(1990).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 777-836.
RX MEDLINE=9417332; PubMed=8127365;
RA Xie X., Harrison D.H., Schlichting I., Sweet R.M., Kalabokis V.N.,
RT Szent-Gyorgyi A.G., Cohen C.;
RL "Structure of the regulatory domain of scallop myosin at 2.8-A
resolution.";
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 777-836.
RX MEDLINE=96419133; PubMed=8805510;
RA Houdusse A., Cohen C.;
RT "Structure of the regulatory domain of scallop myosin at 2-A
resolution: Implications for regulation.";
RL Structure 4:21-32(1996).
CC -!- FUNCTION: Myosin is a protein that binds to F-actin and has ATPase
activity that is activated by F-actin.
CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
and 2 regulatory light chain subunits (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X55714; CAA39247.1; -.
DR PIR; A40997; A40997.
DR PDB; 1SCM; 30-APR-94.
DR PDB; 1WDC; 11-JUL-96.
DR PDB; 1B7T; 12-MAY-99.
DR PDB; 1KK7; 09-OCT-02.
DR PDB; 1KOM; 20-NOV-02.
DR PDB; 1KNO; 20-NOV-02.
DR PDB; 1L20; 30-OCT-02.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; Myosin head.
DR InterPro; IPR004009; Myosin N.
DR InterPro; IPR002928; Myosin tail.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin N; 1.
DR Pfam; PF01576; Myosin tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KM ATP-binding; Calmodulin-binding; 3D-structure.
FT DOMAIN 1 777
FT DOMAIN 778 805 IQ.
FT DOMAIN 836 1938 RODLIKE TAIL (S2 AND LMM DOMAINS).
FT DOMAIN 836 1938 COILED COIL (POTENTIAL).
FT NP_BIND 176 183 ATP (BY SIMILARITY).
FT HELIX 776 822
FT TURN 823 823
FT TURN 825 833
FT TURN 834 836
FT TURN 834 836
SQ SEQUENCE 1938 AA; 22821 MW; ASCCB4127DIA4896 CRC64;
Query Match 8.4%; Score 228; DB 1; Length 1938;
Best Local Similarity 22.4%; Pred. No. 0.033;
Matches 112; Conservative 95; Mismatches 197; Indels 95; Gaps 17;
QY 47 AQSGALRDVSEBLSROLEDI---LSYICVDNNQGGPGGAGCEPAEPEDAERSRTYVAR 103
DB 1272 SQRSLQAENSOLTRQLEDAERHVSVLSEKESQLSSQLEDAERSLEETARSKLQNEVR 1331
QY 104 NCEPEPTPV--VYGEKPSKGDPTBEIRQSDEVGRDRHRRPQEKKAKGLKEITLLMQ 161
DB 1332 NMHADMDAIREQLEEFQESKSDVQRQLSKANNEI---QWRSKFSESGANRTBEL--- 1383
QY 162 TLNLTSTPEKLAALCKKYAELLLEHNSQKMKLLQKKSQVQSKDHLAGEHSAVLA 221
DB 1384 -----EDQKPKLGLKSEAEQTTEAANAKCSALEKAKSLQQE----- 1421
QY 222 RSKLESICRELQHRNRLKE--EGVORA-----REBEKKEVTSHPQVTL 265
DB 1422 ---LEDYSIEVDANASVNMCKKQAFDKTTAEWQAKVNSLQSELENSQKESRGYSAEL 1478
QY 266 NDITQVQHNENRNSKLRQNMELAEKLLKLTQYEL--REEH-IDKVFHKLQQLQVLD 322
DB 1479 YRIKASIEEYQDSIGALRRKNLADEIHLDTQLSSEGSTHELDKARRRLEMEKEELQ 1538
QY 323 AKLQQAQEMLKEAEERHOR-----EKDPLLAEVES-----QRMCELMK--- 361
DB 1539 AALEEAGALEQSEAKVMRAQLFIATVNEIDKRIQEKKEEFNTFRNHQRALESQASL 1598
QY 362 QQTHLKLQQLALYTEKPEPQNTLSKSEVFTTFKQEMEKTKIKK-----LEKET 413
DB 1599 EBAKAGKADAMRIKKKLEQDINELEVALDASNRKAEKTKVKRYQQQIREMQTSIEEQ 1658
QY 414 TMYRSWESSNKALLMAEAK--TVRDKLEGLQVKIQLEKLCALQTERNDLNKRVQDL 472
DB 1659 RQDREARESTN-----MAERCTIMSGEVEELRAALEQAEARAKASDNLADANDRVNEL 1713
QY 473 SAGGQGSLLTDSGPERRPEG 491
DB 1714 TS--QVS-SVQGGKXLEG 1729
RESULT 47
CEP2 HUMAN STANDARD; PRT; 2442 AA.
ID CEP2 HUMAN 014812; O60588; Q9H450;
AC Q9H450; 014812; O60588; Q9H450;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Centrosomal protein 2 (Centrosomal Nex2-associated protein 1) (C-NAP1)
DE (Centrosome protein 250) (Centrosome associated protein CEP250).
GN CEP2 OR CNAP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND AUTOIMMUNE DISEASE.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=98165428; PubMed=9506584;

RA Mack G.J., Rees J., Sandblom O., Balczon R., Fritzler M.J.,
RA Rattner J.B.,
RA "Autoantibodies to a group of centrosomal proteins in human autoimmune
RT sera reactive with the centrosome.";
RL Arthritis Rheum. 41:551-558(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2). INTERACTION WITH NEK2,
RP AND SUBCELLULAR LOCATION DURING THE CELL CYCLE.
RC TISSUE=Placenta;
RX MEDLINE=98311641; PubMed=9647649;
RA Fry A.M., Mayor T., Meraldi P., Stierhof Y.-D., Tanaka K., Nigg E.A.,
RT "C-Nap1, a novel centrosomal coiled-coil protein and candidate
RT substrate of the cell cycle-regulated protein kinase Nek2.";
RL J. Cell Biol. 141:1563-1574(1998).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.S., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leharialho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McCormack L.J., McKay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Beck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Stewart C.A., Sultston J.B.,
RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Symons A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Placenta;
RX MEDLINE=23389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinstead J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krywinski M.I., Skalska U., Smalius D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP PHOSPHORYLATION, AND INTERACTION WITH NEK2 AND PPP1CA.
RX MEDLINE=21060765; PubMed=10880350;
RA Helps N.R., Luo X., Barker H.N., Cohen P.T.W.;
RT "NIMA-related kinase 2 (Nek2), a cell-cycle-regulated protein kinase

RT localized to centrosomes, is complexed to protein phosphatase 1.";
RL Biochem. J. 349:509-518(2000).
RN [6]
RP PHOSPHORYLATION DURING CELL CYCLE.
RX MEDLINE=22135747; PubMed=12140259;
RA Mayor T., Hacker U., Stierhof Y.-D., Nigg E.A.;
RT "The mechanism regulating the dissociation of the centrosomal protein
RT C-Nap1 from mitotic spindle poles.";
RL J. Cell Sci. 115:3275-3284(2002).
CC -!- FUNCTION: Probably plays an important role in centrosome cohesion
CC during interphase.
CC -!- SUBUNIT: Monomer and homodimer (Probable). Forms a complex in
CC vitro with both NEK2 kinase and the PPP1C catalytic subunit of
CC protein phosphatase 1 (PP1).
CC -!- SUBCELLULAR LOCATION: Component of the core centrosome. In
CC interphase cells, it specifically associates with the proximal
CC ends of both mother and daughter centrioles. Associates with the
CC centrosome in interphase cells. In mitotic cells, it dissociates
CC from the mitotic spindle poles. At the end of cell division, it
CC reaccumulates at centrosomes.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q9BV73-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9BV73-2; Sequence=VSP_007372;
CC Name=3;
CC IsoId=Q9BV73-3; Sequence=VSP_007370, VSP_007371;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Ubiquitously and weakly expressed.
CC -!- PTM: Differentially phosphorylated during cell cycle.
CC Phosphorylation may regulate association/dissociation from
CC centrosome. During M phase of mitosis, C-terminal part is
CC phosphorylated by NEK2, suggesting that it may trigger the
CC dissociation from the mitotic centrosome. It is dephosphorylated
CC in vitro by the PP1 phosphatase.
CC -!- DISEASE: Antibodies against CEP2 are present in sera from patients
CC with autoimmune diseases that developed autoantibodies against
CC centrosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF022655; AAC06349.1; -
CC EMBL; AF049105; AAC07988.1; -
CC EMBL; AL121586; CAB89415.1; -
CC EMBL; BC001433; AAH01433.1; -
CC Genes; HGNC:1859; CEP2.
CC Cell cycle; Coiled coil; Phosphorylation; Alternative splicing;
CC Polymorphism.
CC DOMAIN 95 158 COILED COIL (POTENTIAL).
CC DOMAIN 244 352 COILED COIL (POTENTIAL).
CC DOMAIN 395 1172 COILED COIL (POTENTIAL).
CC DOMAIN 1243 2227 COILED COIL (POTENTIAL).
CC DOMAIN 2262 2376 COILED COIL (POTENTIAL).
CC DOMAIN 246 250 POLY-LEU.
CC DOMAIN 464 2171 GLN/LEU-RICH.
CC VARSPLIC 313 313 V -> F (in isoform 3).
CC VARSPLIC 314 2442 /FTId=VSP_007370.
CC VARSPLIC 863 918 Missing (in isoform 3).
CC VARSPLIC 995 995 Missing (in isoform 2).
CC VARIANT 120 120 Q -> H (in dbSNP:2296403).
CC CONFLICT 136 136 L -> I (IN REF. 1).
CC CONFLICT 365 365 E -> A (IN REF. 1).
CC CONFLICT 365 365 H -> L (IN REF. 1).

QY 135 VGDHRRPQEKKAGLGLK-RITLLMOTLNTLSTPEEKLAALCKYAELEHRSOKQ 193
Db 1207 QIDNLRVQKLEKSKBEKLEIDLLSSMSVSKANLEKIKRTLEDQSLRARGKVEE 1266
QY 194 MKLLQKQSOVLVOEKDHLRGEHSKAVLARSKLESCLRELQRHNSLKE--EGVORAESEE 251
Db 1267 ---IQRSLSLQATOKSRLOTEAGELSRQLEBKESIVSOLRSKQAFQTOOTBELKQLEEE 1323
QY 252 EKKKEVTSF-FQVTLNDILOMEQNER-----NSKLR-----QE 285
Db 1324 NKAKNALALALOSSRHDCDLLEQYEBQEGKAEQLQALSANSEVAQWRTKYETDAIOR 1383
QY 286 NMELARLKLQYELPEHIDKY-----FHKD--LQQLVDVAKLQ-QAOEMLKEAE 336
Db 1384 TRELEAKKLAORLQDQEQVEAVNAKCALEKTKQRLQGEVDLMVDVERANSLAAL 1443
QY 337 BRHQEKDPLLEAEVORSOMCSLMKQO--ETHLKOQALYTKFEEFQNTLSKSEVETIF 395
Db 1444 DKQRNFD---KVLAEWTKCBESQABLEASLKSRSLSLTF-KLKNAYEALDQLETV 1499
QY 396 K-----QEMKQTKKI-----KKLEKETTMYRWRSSNKAL----- 427
Db 1500 KRENKYLEOFIADLQEAENGKTIHELEKSKQIELEKADIQALBEAEAALEHEBAKI 1559
QY 428 ---LEMAEKT-----VRDEKLEGLOVKIOR--LEKLCRALQTERNDLNKRVODLSAG 475
Db 1560 LAIQLELTQVKEIDSKIAEKDEEIQKBNVQRTVETVYQSAALDAEVSREAR-LKKX 1618
QY 476 GQGLSTD 482
Db 1619 MEGDLNE 1625

RESULT 50
CENP_HUMAN
ID CENP_HUMAN STANDARD; PRT; 3210 AA.
AC F49454; Q13171; Q13246;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE CENP-F kinetochore protein (Centromere protein P) (Mitosis) (AH
D3 antigen).
GN CENP
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=95348175; PubMed=7542657;
RA Liao H., Winkfein R.J., Mack G., Rattner J.B., Yen T.J.;
RT "CENP-F is a protein of the nuclear matrix that assembles onto
kinetochores at late G2 and is rapidly degraded after mitosis.";
RL J. Cell Biol. 130:507-518 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95379848; PubMed=7651420;
RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
Jones D., Yang-Feng T.L., Lee W.-H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
that is specifically involved in mitotic-phase progression.";
RL Mol. Cell. Biol. 15:5017-5029 (1995).
RN [3]
RP SEQUENCE OF 2194-3210 FROM N.A.
RX MEDLINE=95336446; PubMed=7612011;
RA Li Q., Ye Y., Kapp J.A., Fertig N., Medger T.A. Jr., Joshi H.C.;
RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal
domain sufficient for nuclear localization.";
RL Biochem. Biophys. Res. Commun. 212:220-228 (1995).
RN [4]
RP CHARACTERIZATION.

RX MEDLINE=95370296; PubMed=7642639;
RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
RT "The C terminus of mitotin is essential for its nuclear localization,
centromere/kinetochore targeting, and dimerization.";
RL J. Biol. Chem. 270:19545-19550 (1995).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
interactions with the kinetochore proteins CENP-F and HJURP.";
RL J. Cell Biol. 143:49-63 (1998).
RN [6]
RP FARNESYLATION.
RX MEDLINE=20459117; PubMed=10852915;
RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
Bishop W.R., Kirschmeier P.;
RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E
and CENP-F and alter the association of CENP-E with the
microtubules.";
RL J. Biol. Chem. 275:30451-30457 (2000).
CC -!- FUNCTION: Probably required for kinetochore function, involved in
chromosome segregation during mitosis. Interacts with
retinoblastoma protein (RB), CENP-E and HJURP.
CC -!- SUBUNIT: Homo- or heterodimer.
CC -!- SUBCELLULAR LOCATION: Nuclear matrix (but not in the nucleolus),
reorganization to the kinetochore/centromere (coronal surface of
the outer plate) and the spindle during mitosis.
CC -!- DEVELOPMENTAL STAGE: Gradually accumulates during the cell cycle.
CC -!- PTM: Hyperphosphorylated during mitosis.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EXBL: U19769; AAA82889.1; -
DR EMBL: U30872; AAA82935.1; -
DR EMBL: U25725; AAA86889.1; -
DR PIR: PC4035; PC4035.
DR Genew; HGNC:1857; CENPF.
DR GK; P49454; -
DR MIM; 600236; -
DR GO; GO:0005699; C:kinetochore; TAS.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0005819; C:spindle; TAS.
DR GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR GO; GO:0007088; P:regulation of mitosis; TAS.
KW Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat; Polymorphism;
KW Lipoprotein; Prenylation.
FT DOMAIN 14 197
FT DOMAIN 273 769
FT DOMAIN 823 1328
FT COILED COIL (POTENTIAL).
FT COILED COIL (POTENTIAL).
FT COILED COIL (POTENTIAL).
FT COILED COIL (POTENTIAL).
FT 2 X 177 AA TANDEM REPEATS.
FT 1.
FT 2.
FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT S-farnesyl cysteine.
FT K -> N (in dbSNP:7289).
FT /FTIG=VAR 014839.
FT T -> A (IN REF. 2).
FT L -> Q (IN REF. 2).
FT G -> D (IN REF. 2).
FT MISSING (IN REF. 2).
FT MISSING (IN REF. 2).
FT V -> A (IN REF. 2).
FT V -> L (IN REF. 2).

FT	CONFLICT	2242	2243	ER -> DG (IN REF. 3).
FT	CONFLICT	2335	2335	L -> Q (IN REF. 3).
FT	CONFLICT	2492	2492	D -> N (IN REF. 2).
FT	CONFLICT	2545	2561	ELNVRALHNDQESACK -> SSREWQPCINRTKPVUS (IN REF. 3).
FT	SEQUENCE	3210 AA;	367589 MW;	11D83324960E4334 CRC64;
Query Match				
Best Local Similarity 8.4%; Score 226.5; DB 1; Length 3210;				
Matches 102; Conservative 91; Mismatches 191; Indels 71; Gaps 14;				
Qy	31	GSSQAPKPEGAQARTAQSGALRDVSEELSRQLEDILSTYCVDNNGQGGEDGAGQGPAP	90	
Db	272	GNSSPHLLDQLK--AQQLNRKINLELRL-----QGH--E 305		
Qy	91	PDAEKSTYVARNGEPEPTVWYCEKPSKGDPTNTEBIROSDVGDHRHRRPQSKKAK	150	
Db	306	KEMKGQVNFQELQLOLEKAKVELLEKRVLNKCRDELVRITTAQYDQASTKYTALEQKJK	365	
Qy	151	GIGKEITILMQLNTL-STPEKLAALCKYAELEHRNSQKMKLLQK-----OSQIV 205		
Db	366	KLTEDLSQORNAESARCSLEQKIK--EKEKFEQBELSRQORSFQTLQDECIQMKARLT	422	
Qy	206	QKDHILRGEHKAVLARKLSLCBELQHRNRSLKEEGVQARBEHEEKXSVTSHQVTLT	265	
Db	423	QELQAKNHNVLQAECLKITSVKQOLENNLEEFKQ---KLCRABQAFAS-----QJKE 474		
Qy	266	NDIOLQMBQHNERNRSKLQENNELA-----ERLKKLIQYELREBEHDKVFKHDLQOQ	319	
Db	475	NELRSMEEKKNLLKSHSQKQAREVCHLEABLKNIKQCLNQSONFAEMKAKNTSGB	534	
Qy	320	LVDAKLQAOQEMLKABERHQBKDFLKEAVESQPMC---ELMKQETHLQKQALALYTEK	377	
Db	535	TW---LRDLQEKINOENSLTLEKLVAVADLEKORDCSQDLLKKREHHI-----581		
Qy	378	PEEFONTLSKSSEVFTTPKQEMKMTKKIKLEKETIMVRSWESSNKALLEMASEKTVR	437	
Db	582	-EQLNDKLSKTEKESKALLSALELKKEVEELKBEKTLF-SCKWSENEKLL-----TQM 633		
Qy	438	DKELEGLQVKIORLEKLCRALQTERNDLNKRVQDL	472	
Db	634	ESEKENIASKINHLESTCLKTQOIKGHEYNERNRTL	668	

Search completed: June 7, 2004, 14:38:07
Job time : 36 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:16:35 ; Search time 134 Seconds
(without alignments)
1247.945 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 2702

Sequence: 1 XSSPCQPEAGPEGAQRPSQ.....APSTEASGQTGPQEPSARA 530

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

SPTREML 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2693	99.7	546	4 Q86YW3	Q86YW3 homo sapien
2	2689	99.5	546	4 Q86T86	Q86T86 homo sapien
3	2688	99.5	546	4 Q86T54	Q86T54 homo sapien
4	2677	99.1	546	4 Q86T85	Q86T85 homo sapien
5	1291	47.8	524	11 Q8BHN1	Q8BHN1 mus musculus
6	1283.5	47.5	528	4 Q9NUQ3	Q9NUQ3 homo sapien
7	1266.5	46.9	718	4 Q86T52	Q86T52 homo sapien
8	1265.5	46.8	715	4 Q8N3L3	Q8N3L3 homo sapien
9	1233	45.6	676	13 Q91969	Q91969 gallus galli
10	1223.5	45.5	684	11 Q8VBT1	Q8VBT1 mus musculus
11	1194	44.2	463	11 Q8BUK2	Q8BUK2 mus musculus
12	1008	37.3	505	4 Q8N3S2	Q8N3S2 homo sapien
13	646	23.9	241	11 Q8BP11	Q8BP11 mus musculus
14	598	22.1	116	4 Q86Y86	Q86Y86 homo sapien
15	597.5	22.1	515	5 Q9VBL3	Q9VBL3 drosophila
16	503	18.6	264	5 Q86LQ1	Q86LQ1 branchiostoma

17	462.5	17.1	436	10	Q84VE3	Q84ve3 oryza sativ
18	459	17.0	241	11	Q8BMJ7	Q8bmj7 mus musculus
19	450	16.7	404	10	Q8RXD7	Q8rxd7 arabidopsis
20	442.5	16.4	186	4	Q9POX1	Q9pox1 homo sapien
21	392.5	14.5	150	11	Q8C5K1	Q8c5k1 mus musculus
22	372.5	13.8	335	5	Q22666	Q22666 caenorhabdi
23	369.5	13.7	181	4	Q9BZA4	Q9bza4 homo sapien
24	332	12.3	372	10	Q9FGD8	Q9fgd8 arabidopsis
25	272	10.1	1168	5	Q9VYU0	Q9vyu0 drosophila
26	269	10.0	992	4	Q9NTH6	Q9nth6 homo sapien
27	266.5	9.9	900	3	Q9P3P5	Q9p3p5 neurospora
28	266.5	9.9	1455	4	Q9UPV0	Q9upv0 homo sapien
29	262	9.7	1937	6	Q9TV62	Q9tv62 sus scrofa
30	260	9.6	1941	13	Q9DGM4	Q9dgm4 gallus galli
31	259	9.6	1944	13	Q9DGM5	Q9dgm5 gallus galli
32	257	9.5	1943	13	Q8JG72	Q8jg72 gallus galli
33	256	9.5	1939	13	Q9PTY2	Q9pty2 gallus galli
34	254.5	9.4	609	17	Q8TXA4	Q8txa4 methanopyru
35	253	9.4	1206	11	Q08815	Q08815 rattus norv
36	253	9.4	1940	13	Q8AY28	Q8ay28 gallus galli
37	249.5	9.2	1233	11	Q54988	Q54988 mus musculus
38	249.5	9.2	2139	5	Q07569	Q07569 entamoeba h
39	249	9.2	610	11	Q9CS72	Q9cs72 mus musculus
40	249	9.2	1003	12	Q91LX9	Q91lx9 kaposi's sa
41	248.5	9.2	1005	4	Q9P2I6	Q9p2i6 homo sapien
42	247	9.1	1001	11	Q88664	Q88664 rattus norv
43	246.5	9.1	2055	5	Q8T5C7	Q8t5c7 plasmodium
44	246.5	9.1	2055	5	Q8IHP3	Q8ihp3 plasmodium
45	244	9.0	1202	11	Q9WU41	Q9wu41 mus musculus
46	244	9.0	1212	11	Q8K4T4	Q8k4t4 rattus norv
47	244	9.0	1307	11	Q80U65	Q80u65 mus musculus
48	244	9.0	1391	11	Q922J3	Q922j3 mus musculus
49	243.5	9.0	728	4	Q60561	Q60561 homo sapien
50	243.5	9.0	1001	4	Q9H2K7	Q9h2k7 homo sapien
51	243.5	9.0	1598	11	Q922D2	Q922d2 mus musculus
52	242	9.0	1937	6	Q8MJV1	Q8mjv1 equus cabal
53	242	9.0	1938	6	Q9GUP9	Q9gup9 oryctolagus
54	242	9.0	1939	6	Q9TV63	Q9tv63 sus scrofa
55	242	9.0	1939	6	Q9TV61	Q9tv61 sus scrofa
56	241.5	8.9	1089	12	Q40947	Q40947 kaposi's sa
57	241.5	8.9	1508	5	Q8GJ39	Q8gj39 dictyosteli
58	241.5	8.9	1864	5	Q8SNQ7	Q8snq7 loligo peal
59	241	8.9	1001	13	Q7ZIG0	Q7zj0 xenopus lae
60	240.5	8.9	2033	10	Q7XKH4	Q7xeh4 oryza sativ
61	240	8.9	879	13	Q9YHD8	Q9yhd8 rana catesb
62	240	8.9	1931	13	Q91973	Q91973 coturnix co
63	240	8.9	1938	6	Q8MJV0	Q8mjv0 equus cabal
64	239.5	8.9	901	11	Q7TN77	Q7tn77 mus musculus
65	239.5	8.9	1197	3	Q9C3Y7	Q9c3y7 candida alb
66	239.5	8.9	1201	11	Q8CIA8	Q8cia8 mus musculus
67	239.5	8.9	1413	11	Q8CGB3	Q8cgb3 mus musculus
68	239	8.8	1057	5	Q8MRS3	Q8mrs3 drosophila
69	239	8.8	1938	6	Q9BE40	Q9be40 bos taurus
70	239	8.8	1940	6	Q9BE41	Q9be41 bos taurus
71	239	8.8	2115	4	Q14980	Q14980 homo sapien
72	238.5	8.8	1935	6	Q9BE39	Q9be39 bos taurus
73	238	8.8	892	11	Q63939	Q63939 rattus sp.
74	238	8.8	1036	12	Q9DUM3	Q9dum3 kaposi's sa
75	237.5	8.8	1177	4	Q86TC3	Q86tc3 homo sapien
76	237.5	8.8	1213	4	Q727B0	Q727b0 homo sapien
77	237	8.8	631	13	Q7Z219	Q7z219 brachydanio
78	237	8.8	915	5	Q86SR0	Q86sr0 dictyosteli
79	236	8.7	970	5	Q9N8Y9	Q9n8y9 trypanosoma
80	236	8.7	1935	6	Q8MJU9	Q8mjju9 equus cabal
81	236	8.7	1935	11	Q91283	Q91283 mus musculus
82	236	8.7	1941	13	Q8UWA0	Q8uwa0 gallus galli
83	235.5	8.7	1437	11	Q80XQ1	Q80xq1 mus musculus
84	235	8.7	728	11	Q8K454	Q8k454 mus musculus
85	235	8.7	1229	5	Q9N322	Q9n322 aequipecten
86	235	8.7	1243	5	Q9N321	Q9n321 aequipecten
87	235	8.7	1253	5	Q9NU20	Q9nu20 aequipecten
88	235	8.7	1951	5	Q17042	Q17042 aequipecten
89	234.5	8.7	650	4	Q8N9W4	Q8n9w4 homo sapien

90	234.5	8.7	1231	11	O55092	O55092 cavia porce
91	234.5	8.7	1235	4	O9H2G2	O9h2g2 homo sapien
92	234.5	8.7	1596	5	O8I044	O8i044 plasmodium
93	234.5	8.7	1935	6	O9GKR1	O9gkri sus scrofa
94	234	8.7	891	13	O90WH5	O90wh5 gallus gall
95	234	8.7	976	12	O9DUN0	O9dun0 kaposi's sa
96	234	8.7	1931	13	O910C5	O910c5 gallus gall
97	234	8.7	1935	4	O9H1D5	O9h1d5 homo sapien
98	233.5	8.6	1512	4	O8N8B9	O8n8b9 homo sapien
99	233.5	8.6	1940	5	O9I1F6	O9i1f6 plasmodium
100	233.5	8.6	789	5	O9U7E3	O9u7e3 pecten maxi
101	233	8.6	1941	4	O8E6AC9	O8e6ac9 dictyosteli
102	233	8.6	1941	4	O8E6T56	O8e6t56 homo sapien
103	233	8.6	1950	5	O26080	O26080 placopecten
104	233	8.6	2760	5	O8I5Y2	O8i5y2 plasmodium
105	232.5	8.6	1432	10	O23230	O23230 arabidopsis
106	232	8.6	2094	11	O8O335	O8o335 mus musculus
107	231.5	8.6	1129	12	O9QK71	O9qk71 kaposi's sa
108	231.5	8.6	1927	5	O25142	O25142 halocynthia
109	231	8.5	1388	11	P70336	P70336 mus musculus
110	231	8.5	1833	11	O80T68	O80t68 mus musculus
111	231	8.5	2148	5	O9BJD3	O9bjd3 physarum po
112	230.5	8.5	917	11	O8CJC0	O8cjc0 mus musculus
113	230.5	8.5	1088	4	O8IUD3	O8iud3 homo sapien
114	230.5	8.5	1116	4	O8IUD2	O8iud2 homo sapien
115	230.5	8.5	1363	3	O874Y4	O874y4 podospora a
116	230.5	8.5	1935	5	O44934	O44934 loligo peal
117	230.5	8.5	3187	11	O63714	O63714 rattus norv
118	229.5	8.5	1941	5	O26079	O26079 placopecten
119	229	8.5	611	4	O14905	O14905 homo sapien
120	229	8.5	1193	4	O7Z757	O7z757 homo sapien
121	229	8.5	1305	10	O9FJ35	O9fj35 arabidopsis
122	228.5	8.5	122	13	O9YHD7	O9ynd7 rana catesb
123	228.5	8.5	1006	4	O9E157	O9e157 homo sapien
124	228.5	8.5	1889	4	O9H430	O9h430 homo sapien
125	228.5	8.5	1932	13	O98TQ4	O98tq4 notothenia
126	228.5	8.5	2010	4	O9P216	O9p216 homo sapien
127	228	8.4	910	13	O7T2F8	O7t2f8 brachydanio
128	228	8.4	948	11	O8CIY9	O8ciy9 rattus norv
129	228	8.4	948	11	O8I1U3	O8i1u3 rattus norv
130	228	8.4	995	5	O9N7J9	O9nj9 aequipecten
131	228	8.4	1205	5	P92021	P92021 caenorhabdi
132	228	8.4	1219	5	O9N7J3	O9nj23 aequipecten
133	228	8.4	1388	6	O28021	O28021 bos taurus
134	227.5	8.4	1162	12	O98148	O98148 kaposi's sa
135	227.5	8.4	1744	10	O9SA62	O9sa62 arabidopsis
136	227.5	8.4	1937	13	O9IBD4	O9ibd4 gallus gall
137	227.5	8.4	1945	5	O98LDO	O9bld0 patinopecte
138	227	8.4	585	11	O8CAU2	O8cauz mus musculus
139	227	8.4	948	4	O9UIK7	O9uik7 homo sapien
140	227	8.4	992	4	O8IUD5	O8iud5 homo sapien
141	227	8.4	1003	4	O9UP51	O9ups1 homo sapien
142	227	8.4	1132	5	O8SZW3	O8szw3 drosophila
143	226.5	8.4	688	4	O96NL6	O96nl6 homo sapien
144	226.5	8.4	1140	4	O9ULE5	O9ule5 homo sapien
145	226.5	8.4	1790	3	O07380	O07380 saccharomyc
146	226	8.4	891	5	O18430	O18430 geodia cydo
147	226	8.4	908	5	O8INC2	O8inc2 drosophila
148	226	8.4	920	5	O9IH49	O9ih49 drosophila
149	226	8.4	963	4	O96HNS	O9ehn5 homo sapien
150	226	8.4	1134	5	O9INC1	O9inc1 drosophila

ALIGNMENTS

RESULT 1					
O86YW3	PRELIMINARY;	PRT;	546 AA.		
ID	O86YW3				
AC	O86YW3;				
DT	01-JUN-2003	(Tremblrel. 24, Created)			
DT	01-JUN-2003	(Tremblrel. 24, Last sequence update)			
DT	01-JUN-2003	(Tremblrel. 24, Last annotation update)			

DE	Taxilin.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RC	SEQUENCE FROM N.A.				
RP	TISSUE=Brain;				
RX	MEDLINE=22448667; PubMed=12558796;				
RA	Nogami S., Satoh S., Nakano M., Shimizu H., Fukushima H., Maruyama A.,				
RA	Terano A., Shirataki H.				
RT	"taxilin; a novel syntaxin-binding protein that is involved in Ca2+-				
RT	dependent exocytosis in neuroendocrine cells.";				
RL	Genes Cells 8:17-28(2003).				
DR	EMBL; AF516206; AAO42465.1; ..				
SQ	SEQUENCE 546 AA; 61891 MW; 698CD74F78897DF6 CRC64;				
Query Match 99.7%; Score 2693; DB 4; Length 546;					
Best Local Similarity 99.8%; Pred. No. 1.1e-120; Indels 0; Gaps 0;					
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
QY	1	KSSPGQPEAGPEGAQERSQAAPAVEAGPGSSQAPRPEGAQARTAQSGALRDVSEEL	60		
Db	17	KSSPGQPEAGPEGAQERSQAAPAVEAGPGSSQAPRPEGAQARTAQSGALRDVSEEL	76		
QY	61	ROLEDILTSTCYDNNQGGEDGAGQEPAPDAEKSRITYVARNGBEPETPVVYGEKBP	120		
Db	77	ROLEDILTSTCYDNNQGGEDGAGQEPAPDAEKSRITYVARNGBEPETPVVYGEKBP	136		
QY	121	KDPTNTEIRQSDVGDVDRDHRPQEKKAAGLKGKEITLLMOTINTLSTPEEKLAALCKY	180		
Db	137	KDPTNTEIRQSDVGDVDRDHRPQEKKAAGLKGKEITLLMOTINTLSTPEEKLAALCKY	196		
QY	181	AELLEHRNSQOMKLLQKQOLQVQKDLHGEHSKAVLARSKLESCLRELQHRNSLK	240		
Db	197	AELLEHRNSQOMKLLQKQOLQVQKDLHGEHSKAVLARSKLESCLRELQHRNSLK	256		
QY	241	EEGVQARBEERKEKRVTSHFQVTLNDILOMEQHNRNSKLRQENNELAERLKKI	300		
Db	257	EEGVQARBEERKEKRVTSHFQVTLNDILOMEQHNRNSKLRQENNELAERLKKI	316		
QY	301	ELBEEHIDKVFHKLOLQOLVDKLOQAEMLKEABERHOREKDFLLKAVESQRMCELM	360		
Db	317	ELBEEHIDKVFHKLOLQOLVDKLOQAEMLKEABERHOREKDFLLKAVESQRMCELM	376		
QY	361	KQETHLKKOALALYTEKFEFQNTLSKSEVFTTFKQEKEMTKTKIKKLEKTTMYRSRW	420		
Db	377	KQETHLKKOALALYTEKFEFQNTLSKSEVFTTFKQEKEMTKTKIKKLEKTTMYRSRW	436		
QY	421	ESSNKALLEMAEKTVRDKSEGLQVKIQRLEKLCALOTERNDLNKRVDLSAGGGSL	480		
Db	437	ESSNKALLEMAEKTVRDKSEGLQVKIQRLEKLCALOTERNDLNKRVDLSAGGGSL	496		
QY	481	TDGSPERRPBGQAQAPSSPRVTEAPCYCAPSTEASGQTGPQPTSARA	530		
Db	497	TDGSPERRPBGQAQAPSSPRVTEAPCYCAPSTEASGQTGPQPTSARA	546		
RESULT 2					
O86T86	PRELIMINARY;	PRT;	546 AA.		
ID	O86T86				
AC	O86T86;				
DT	01-JUN-2003	(Tremblrel. 24, Created)			
DT	01-JUN-2003	(Tremblrel. 24, Last sequence update)			
DT	01-JUN-2003	(Tremblrel. 24, Last annotation update)			
DE	Hypothetical protein DKFZp451I0918.				
GN	DKFZP451I0918.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				

RA Ansoorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.,
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AU832636; CAD89951.1; -;
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61892 MW; FEBCAD4F753F7FF1 CRC64;

Query Match 99.5%; Score 2689; DB 4; Length 546;
Best Local Similarity 99.6%; Pred. No. 1.6e-120;
Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAGSALRDVSEELS 60
DB 17 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAGSALRDVSEELS 76
QY 61 RQLEDILSTYCVNNQGGPGEAGQEPAPEDAEKSRITYVARNGEPPTPVVYGEKPS 120
DB 77 RQLEDILSTYCVNNQGGPGEAGQEPAPEDAEKSRITYVARNGEPPTPVVYGEKPS 136
QY 121 KGPNTTEIRQSDVGDGRDHRPQSKKAGLKGKITTLLMOTLNTLSTPEEKLAALCKKY 180
DB 137 KGPNTTEIRQSDVGDGRDHRPQSKKAGLKGKITTLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 AELEEHRSQKMKLLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSILK 240
DB 197 AELEEHRSQKMKLLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSILK 256
QY 241 EGVQARBEERKREKVTSHFQVTLNDIQLQMEQHNERNKSLRQNMELAEELKXLIBQY 300
DB 257 EGVQARBEERKREKVTSHFQVTLNDIQLQMEQHNERNKSLRQNMELAEELKXLIBQY 316
QY 301 ELREEHIDKVFHKLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSILK 360
DB 317 ELREEHIDKVFHKLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSILK 376
QY 361 KQETHLQKQALYTKPEEFQNTLSKSEVFTTTPQEMEKMTKKIKKLEKTTMYRSRW 420
DB 377 KQETHLQKQALYTKPEEFQNTLSKSEVFTTTPQEMEKMTKKIKKLEKTTMYRSRW 436
QY 421 ESSNKALLEMAEKTVRDKEGLQVQKDLHGEHSKAVLARSKLESICRELQRNRSILK 480
DB 437 ESSNKALLEMAEKTVRDKEGLQVQKDLHGEHSKAVLARSKLESICRELQRNRSILK 496
QY 481 TDSGPRRPEGGAQAPSPRVTEAPCPVGPAPSTASGOTGPQPTTSARA 530
DB 497 TDSGPRRPEGGAQAPSPRVTEAPCPVGPAPSTASGOTGPQPTTSARA 546

RESULT 3

Q86T54 PRELIMINARY; PRT; 546 AA.
ID Q86T54
AC Q86T54;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Hypothetical protein DKZP451K215.
GN DKZP451K215.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
RA Koehrer K., Meyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.,
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AU832338; CAD91138.1; -;
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61861 MW; 728D674F78897DF6 CRC64;
Query Match 99.5%; Score 2688; DB 4; Length 546;
Best Local Similarity 99.6%; Pred. No. 1.8e-120;

Matches 528; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAGSALRDVSEELS 60
DB 17 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAGSALRDVSEELS 76
QY 61 RQLEDILSTYCVNNQGGPGEAGQEPAPEDAEKSRITYVARNGEPPTPVVYGEKPS 120
DB 77 RQLEDILSTYCVNNQGGPGEAGQEPAPEDAEKSRITYVARNGEPPTPVVYGEKPS 136
QY 121 KGPNTTEIRQSDVGDGRDHRPQSKKAGLKGKITTLLMOTLNTLSTPEEKLAALCKKY 180
DB 137 KGPNTTEIRQSDVGDGRDHRPQSKKAGLKGKITTLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 AELEEHRSQKMKLLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSILK 240
DB 197 AELEEHRSQKMKLLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSILK 256
QY 241 EGVQARBEERKREKVTSHFQVTLNDIQLQMEQHNERNKSLRQNMELAEELKXLIBQY 300
DB 257 EGVQARBEERKREKVTSHFQVTLNDIQLQMEQHNERNKSLRQNMELAEELKXLIBQY 316
QY 301 ELREEHIDKVFHKLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSILK 360
DB 317 ELREEHIDKVFHKLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSILK 376
QY 361 KQETHLQKQALYTKPEEFQNTLSKSEVFTTTPQEMEKMTKKIKKLEKTTMYRSRW 420
DB 377 KQETHLQKQALYTKPEEFQNTLSKSEVFTTTPQEMEKMTKKIKKLEKTTMYRSRW 436
QY 421 ESSNKALLEMAEKTVRDKEGLQVQKDLHGEHSKAVLARSKLESICRELQRNRSILK 480
DB 437 ESSNKALLEMAEKTVRDKEGLQVQKDLHGEHSKAVLARSKLESICRELQRNRSILK 496
QY 481 TDSGPRRPEGGAQAPSPRVTEAPCPVGPAPSTASGOTGPQPTTSARA 530
DB 497 TDSGPRRPEGGAQAPSPRVTEAPCPVGPAPSTASGOTGPQPTTSARA 546

RESULT 4

Q86T85 PRELIMINARY; PRT; 546 AA.
ID Q86T85
AC Q86T85;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Hypothetical protein DKZP451J0118.
GN DKZP451J0118
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Ansoorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.,
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AU832637; CAD89952.1; -;
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61846 MW; 36718BAE3AA7B6C2 CRC64;
Query Match 99.1%; Score 2677; DB 4; Length 546;
Best Local Similarity 99.2%; Pred. No. 6.1e-120;
Matches 526; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAGSALRDVSEELS 60
DB 17 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAGSALRDVSEELS 76
QY 61 RQLEDILSTYCVNNQGGPGEAGQEPAPEDAEKSRITYVARNGEPPTPVVYGEKPS 120
DB 77 RQLEDILSTYCVNNQGGPGEAGQEPAPEDAEKSRITYVARNGEPPTPVVYGEKPS 136


```
QY 250 EEEKREKVTSHFQVTLNDIOLQWQHNRNSKLRQENMLAERLKLHBYHLREHIDK 309
Db 233 EBERREKETAHFQITLDEIQAOLQEHDIENAKLRQENIELGKXKLHBYALREHIDK 292
QY 310 VFHKLQOOLVDKLOQAOQEMLEAEERHOREKDFLLKEAVESQRMCLAKQOETHLKQ 369
Db 293 VFHKLQOOLVDKLOQAOQEMLEAEERHOREKDFLLKEAVESQRMCLAKQOETHLKQ 352
QY 370 QALYTEKFEFQNTLSKSEVFTTFKQEBEYKTKKLEKHTMYRWRWSSNKALLR 429
Db 353 QLSLYMDKFEFQNTLSKSEVFTTFKQEBEYKTKKLEKHTMYRWRWSSNKALLR 412
QY 430 MABEKTVRKELGLOVKIQRLEKLCRALQTERNDLNKRVQ-----DL 472
Db 413 MABEKTVRKELGLOVKIQRLEKLCRALQTERNDLNKRVQ-----DL 472
QY 473 SAGQGSGLT--DSGPE-----RRPEGGAQA-PSSPR--VTEAPCPGAPSTEA 516
Db 473 ATPVMPCTALDSHKSINTSSKALGAHLEAPKQORSQSAVQKPPSTGSAPEIES 526
RESULT 7
ID Q86T52 PRELIMINARY; PRT; 718 AA.
AC Q86T52;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein DKFZp451G083 (Fragment).
GN DKFZP451G083.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Oeanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL819198; CAD91140.1; -.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 718 AA; 80290 MW; 16C3BF22C19559A4 CRC64;
Query Match 46.9%; Score 1266.5; DB 4; Length 718;
Best Local Similarity 46.4%; Pred. No. 9.6e-53;
Matches 276; Conservative 90; Mismatches 134; Indels 95; Gaps 10;
QY 18 PSQAAPAVEAR-----GPG-SSQAP-----RKPEGAQART-----AQSGLR 53
Db 28 PSDTTPCKXANHSQLSAERQSTPFGDSSSLPSHNGLEKEDGQDSPTVPVQPEKASVHP 87
QY 54 DVSELSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYVARNGEPPTPVV 113
Db 88 DISELNARQLEDIINTY--GSAASTAGKESASARASEQENAESPDN---EDGDCERTTEE 142
QY 114 YGEKEPSKGDPTNTEELRQSDVGRDHRHPQEKKAQGLGKEITLLMQTLNTLSTPEKL 173
Db 143 AGREPVASGEPPTVK---EPVSNKEQK---LEKQILGKGLKEANLLMQNLKQTPPEKF 196
QY 174 AALCKKYAELLBEHRNSQOMKLLQKQSOVLQEKDHLRGHSGKAVLARSKLSLCRELQ 233
Db 197 DFLFKKYAELLDEHRTQKLLQKQVQIQEKDQLOQGEHSRAILARSKLSLCRELQ 256
QY 234 RHNRSLKEGVQVQARBEERKKEVTSHTFQVTLNDIOLQWQHNRNSKLRQENMLAERL 293
Db 257 RHNRSLKEGVQVQARBEERKKEVTSHTFQVTLNDIOLQWQHNRNSKLRQENMLAERL 316
QY 294 KCLLEQYELREHIDKVFHKLQOOLVDKLOQAOQEMLEAEERHOREKDFLLKEAVES 353
Db 317 KSIIDQYELREHIDKVFHKLQOOLVDKLOQAOQEMLEAEERHOREKDFLLKEAVES 376
```

```
QY 354 QRMCELMKQOETHLKOOLALYTEKFEFQNTLSKSEVFTTFKQEBEYKTKKLEKHT 413
Db 377 KLOAKVLEQOETHLQOALYTEKFEFQNTLSKSEVFTTFKQEBEYKTKKLEKHT 436
QY 414 TMYRSRWSSNKALLREAEKRYTVDKLEGLQVKIQRLEKLCRALQTERNDLNKEVDLS 473
Db 437 ATWKARFENCNKALIDMTERRKALAKKEYECFVMKIGRLNLCRALQOBERNEUHKIRDAE 496
QY 474 AGGQGSGLTDSGPERPEG----- 491
Db 497 ISEKDDQSHNSDEEPESNVSDQIDEAEVNSVQTAVKNLATAFMIIHHPSTPHQSKE 556
QY 492 -----PQAQAPSSP-----RVTEAPCPGAPSTEAQSGTQFQBPSTA 528
Db 557 TQPEIGSSQESADAALKEPQPLIPSRDSSEPLPLTPQAEAEQSGDAEPESKA 611
RESULT 8
ID Q8N3L3 PRELIMINARY; PRT; 715 AA.
AC Q8N3L3;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein DKFZp451P022 (Fragment).
GN DKFZP451P022.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Oeanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL834248; CAD38924.2; -.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 715 AA; 79798 MW; 11C50191BCD26582 CRC64;
Query Match 46.8%; Score 1265.5; DB 4; Length 715;
Best Local Similarity 46.4%; Pred. No. 1.1e-52;
Matches 276; Conservative 90; Mismatches 134; Indels 95; Gaps 10;
QY 18 PSQAAPAVEAR-----GPG-SSQAP-----RKPEGAQART-----AQSGLR 53
Db 25 PSDTTPCKXANHSQLSAERQSTPFGDSSSLPSHNGLEKEDGQDSPTVPVQPEKASVHP 84
QY 54 DVSELSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYVARNGEPPTPVV 113
Db 85 DISELNARQLEDIINTY--GSAASTAGKESASARASEQENAESPDN---EDGDCERTTEE 139
QY 114 YGEKEPSKGDPTNTEELRQSDVGRDHRHPQEKKAQGLGKEITLLMQTLNTLSTPEKL 173
Db 140 AGREPVASGEPPTVK---EPVSNKEQK---LEKQILGKGLKEANLLMQNLKQTPPEKF 193
QY 174 AALCKKYAELLBEHRNSQOMKLLQKQSOVLQEKDHLRGHSGKAVLARSKLSLCRELQ 233
Db 194 DFLFKKYAELLDEHRTQKLLQKQVQIQEKDQLOQGEHSRAILARSKLSLCRELQ 253
QY 234 RHNRSLKEGVQVQARBEERKKEVTSHTFQVTLNDIOLQWQHNRNSKLRQENMLAERL 293
Db 254 RHNRSLKEGVQVQARBEERKKEVTSHTFQVTLNDIOLQWQHNRNSKLRQENMLAERL 313
QY 294 KCLLEQYELREHIDKVFHKLQOOLVDKLOQAOQEMLEAEERHOREKDFLLKEAVES 353
Db 314 KSIIDQYELREHIDKVFHKLQOOLVDKLOQAOQEMLEAEERHOREKDFLLKEAVES 373
QY 354 QRMCELMKQOETHLKOOLALYTEKFEFQNTLSKSEVFTTFKQEBEYKTKKLEKHT 413
Db 374 KLOAKVLEQOETHLQOALYTEKFEFQNTLSKSEVFTTFKQEBEYKTKKLEKHT 433
```



```
RESULT 11
Q8BUK2 ID Q8BUK2 PRELIMINARY; PRT; 463 AA.
AC DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Muscle-derived protein XBP77 variant 1 (Fragment).
GN 231000IN14RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.",
RL Nature 420:563-573(2002).
DR EMBL; AK084639; BAC39238.1; -.
DR MGD; MGI:1916756; 231000IN14RIK.
FT NON TER 463
SQ SEQUENCE 463 AA; 54090 MW; 18E9F881D73D59AA CRC64;

Query Match 44.2%; Score 1194; DB 11; Length 463;
Best Local Similarity 52.7%; Pred. No. 1.7e-49;
Matches 251; Conservative 72; Mismatches 103; Indels 50; Gaps 6;

QY 18 PSQAFAVAEAGPGSQAPKPEGAQARTASQAL-----RDVSELSRQLSDILSTYC 71
DB 16 PPGDSSLSNQPGKQDGCSTSGA-PEQGSLLHPEKGAHVABELSRQLSDIISTY- 73
QY 72 VDNNOGPGEDGAGQEPAPEDAEKSTYVARNGEPEPTPVVYGEKPSKGDPTESIRQ 131
DB 74 -----GSAASPRGKSTSETKEQPPNTEAPNEDVDYEE-----TTSEI-- 112
QY 132 SDEVDGRDHRPQ-----EKKKAGLKGKEITLLMQTLNTLTSPKLA 175
DB 113 -----DREPTAPEAPAAKEPVSKQEKLEKILKGLKEANLLQNLNKLQAPSEKLD 167
QY 176 LCKYAELEBEHNSQKMLQKQSQOLVQEKDHLRGESKAVLARSKLSLCRELQSH 235
DB 168 LFKKYTELDEHRTQCKLKLKQQAQTOREKQDQSEHNRVAVLARSKLSLCRELQSH 227
QY 236 NRSIKKEGVORAREEERKEVTSHTFOVTLNDIQLQMEQHNRNSKLRQENMELAEELKX 295
DB 228 NKILKETLQAREEERKEVTSHTFOVTLNDIQLQMEQHNRNSKLRQENMELAEELKX 287
QY 296 LIEQYELREHIDKVFHKLQOLVDKIQQAQEMLKZAEERHOREKDFLLKAEVSR 355
DB 288 IIDQYELREHIDKVFHKLQOLVDKIQQAQEMLKZAEERHOREKDFLLKAEVSR 347
QY 356 MCELAKQOETHLQKQALYKTEKEFEFNTLSKSEVFTTQKQEMKVTKKIKLEKETT 415
DB 348 QAKVLKQETVLOQLTLYSGRFEFQSTLTKSNEVFATKQEMDKTYYKKKLEKDTAT 407
QY 416 YRSRWSSNKLALMELAEKTVRDKELGLQVKTORLEKLCRALQTERNDLNKRVQD 471
DB 408 WKARFENCNALLDMIEKALRAKEVECFVWKIQRLENLCRALQTERNDLNKRVQD 463

RESULT 12
Q8N3S2 ID Q8N3S2 PRELIMINARY; PRT; 505 AA.
AC DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP451A175.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.",
RL Nature 420:563-573(2002).
DR EMBL; AK078477; BAC37296.1; -.
DR MGD; MGI:1194910; Rbbp7.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0018564; P:transcriptional repressor activity; IDA.
DR GO; GO:0000122; P:negative regulation of transcription from P...; IDA.
DR GO; GO:0000122; P:negative regulation of transcription from P...; IDA.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.",
RL Nature 420:563-573(2002).
DR EMBL; AK078477; BAC37296.1; -.
DR MGD; MGI:1194910; Rbbp7.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0018564; P:transcriptional repressor activity; IDA.
DR GO; GO:0000122; P:negative regulation of transcription from P...; IDA.
DR GO; GO:0000122; P:negative regulation of transcription from P...; IDA.

Query Match 37.3%; Score 1008; DB 4; Length 505;
Best Local Similarity 51.0%; Pred. No. 1.3e-40;
Matches 203; Conservative 64; Mismatches 71; Indels 60; Gaps 2;

QY 191 QKMKLLQKQSQOLVQEKDHLRGESKAVLARSKLSLCRELQSHNRSLKKEGVORAREE 250
DB 1 QKMKLLQKQSQOLVQEKDHLRGESKAVLARSKLSLCRELQSHNRSLKKEGVORAREE 60
QY 251 BEKREKVTSHFOVTLNDIQLQMEQHNRNSKLRQENMELAEELKXIKLEKETTMYR 310
DB 61 BEKREKVTSHFOVTLNDIQLQMEQHNRNSKLRQENMELAEELKXIKLEKETTMYR 120
QY 311 FKHKLQOLVDKIQQAQEMLKZAEERHOREKDFLLKAEVSRQENMELAEELKXIK 370
DB 121 FKHKLQOLVDKIQQAQEMLKZAEERHOREKDFLLKAEVSRQENMELAEELKXIK 180
QY 371 LALYKTEKEFEFNTLSKSEVFTTQKQEMKVTKKIKLEKETTMYR 430
DB 181 LALYKTEKEFEFNTLSKSEVFTTQKQEMKVTKKIKLEKETTMYR 240
QY 431 ABEKTVRDKELGLQVKTORLEKLCRALQTERNDLNKRVQDLSAGQSLTDSGPERRPE 490
DB 241 IBEKALRAKEVECFVWKIQRLENLCRALQTERNDLNKRVQDLSAGQSLTDSGPERR 300
QY 491 G-----PGAQ 495
DB 301 SNVSVQDEIDAEVNSVQTVAKLATATFMIHHPSTHQSKETQPEIGSSQESADA 360
QY 496 APSSP-----RVTEAPCYPGAPSTEAAGTGPQPTSA 528
DB 361 EPEQPLPSRDSSEPLPLTPQABAGSGSDAEPSSKA 398

RESULT 13
Q8BP11 ID Q8BP11 PRELIMINARY; PRT; 241 AA.
AC DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RBBP7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mullerian duct includes surrounding region;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.",
RL Nature 420:563-573(2002).
DR EMBL; AK078477; BAC37296.1; -.
DR MGD; MGI:1194910; Rbbp7.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0018564; P:transcriptional repressor activity; IDA.
DR GO; GO:0000122; P:negative regulation of transcription from P...; IDA.
DR GO; GO:0000122; P:negative regulation of transcription from P...; IDA.
```


Db 106 KNESSLQIKVEERRKESQTKFSSSLNDVQKSLAKNNEENIKLRDYNIEWTXKLLKLABQ 165
Qy 300 YELRESHIDKVPKHDLQOOLVDAKLOQAOEMLKEAERHQRKDFLLKEAVESQRMCEL 359
Db 166 YQTRSHOLEKLNQVQLEAHLQAKQKQCVZAMKEILSKENQIGLEKLMQOQRAIKD 225
Qy 360 MKQOETHLKOALYTKFEPFNTLSKSSEVFTTFKQENBOMTKKIKLEKFTMYRSR 419
Db 226 LTRDREHLQKELNITAKYDDFQSLQKSNVFGSVKVELEKXSKTKTKIERKALGWROK 285
Qy 420 WESSNKALLEMAEKTVRDKELHGLQVKTQRIKLCRALQYTERNDINKRVOD 471
Db 286 YEKANAMVIDLATEKSLQTHQSERLQKIQOOLKJLRAQLQERTTLHKLRD 337

RESULT 16
Q86LQ1 PRELIMINARY; PRT; 264 AA.
AC Q86LQ1 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical muscle-derived protein.
OS Branchiostoma belcheri (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7741;
RN [1]
R2 SEQUENCE FROM N.A.
RA Zhang H., Lin Y.
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY191780; AA045170.1; -.
KW Hypothetical protein.
SQ SEQUENCE 264 AA; 31199 MW; EA0B82C96708815B CRC64;

Query Match 18.6%; Score 503; DB 5; Length 264;
Best Local Similarity 51.2%; Pred. No. 7.2e-17;
Matches 103; Conservative 33; Mismatches 57; Indels 8; Gaps 1;
Qy 282 LRQNMELAEKLIROYELREHIDKVPKHDLQOOLVDAKLOQAOEMLKEAERHQR 341
Db 1 MKEENMELCGKLSLVEQYERKEEVEKLNKQDLERQLAEAKMQQATMLABEKEFLT 60
Qy 342 EKDFLLKEAVESQRMCELMKQETHLKOALYTKFEPFNTLSKSSEVFTTFKQEMK 401
Db 61 EKQVLLKENYAHRSFQMAIQENMKMQLKLYTEKFEELHKLNQSNVFNKFEVMDT 120
Qy 402 MTKKIKLEKFTMYRSRWSSESKALLEMAEKTVRDKELHGLQVKTQRIKLCRALQTE 461
Db 121 MKRMKLEKESFQWRTKWEKSNVTLAWASEKQTRDKELIMLRKTCGKLENLCRALQNA 180
Qy 462 RNDINKRVODLSAGQSLTD 482
Db 181 RSD-----TEGGADKVT 193

RESULT 17
Q84VE3 PRELIMINARY; PRT; 436 AA.
AC Q84VE3 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Muscle derived-like protein.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eubartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
R2 SEQUENCE FROM N.A.
RA STRAIN-cv. Nipponbare;
RC Cooper B., Hutchinson D., Park S., Guimil S., Luginbuhl P., Ellero C.,

RA Goff S., Glazebrook J.;
RT "Identification of rice (Oryza sativa) Proteins Linked to the Cyclin-
Mediated Regulation of the Cell Cycle."
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY224541; AA072661.1; -.
SQ SEQUENCE 436 AA; 48148 MW; 91B0958FB88E2328 CRC64;

Query Match 17.1%; Score 462.5; DB 10; Length 436;
Best Local Similarity 28.9%; Pred. No. 1e-14;
Matches 140; Conservative 83; Mismatches 156; Indels 105; Gaps 15;
Qy 12 EGAQRPSCQAPAVEA-----EGSGSQAPKPSGAQARTAQSGALRDVSELSQLED 65
Db 2 EGS---PATRLPEADSLPDGFFVSSAADAQAPPP-----ADAGPASES---- 42
Qy 66 ILSTYCVNNQGGPGEDGAQGEPA-----EPEDAERSRTYVARGPPE---- 108
Db 43 -LGLQADATVGGGGDETLGAPPSTLASVAQDTLDAYSADALQSLTVGSSAABPERAL 101
Qy 109 -----PTPVYGEKPSKDPNTEIRQSDVEGDRDRHREPOEKKKAKGLKGTLLM 160
Db 102 GEPVADAGAVPVVADAKESK--ESSVVEQVESMAD-----QKVITAEKGG-- 145
Qy 161 QTLNLTSTPEEKLAALCKKYABLLBEHRNSQOMKLLQKKQ-SQLVQEKDHLRGEHSAV 219
Db 146 -----EQKRVKVKSVKEDRELFEALQAYHKVVAERDAI 181
Qy 220 LARSKLESICRELQHNRLKEEGVQORARBEERKEKEVTISHFQVTLNDIQLQK-EQHN 278
Db 182 AVKEKLESICREBFQONKALKEECRRVSTEGQMMELESDKFNNAIKDVSXLDQKNEC 241
Qy 279 NSKLRQNMELAEKLIROYELREHIDKVPKHDLQOOLVDAKLOQAOEMLKEAER 338
Db 242 IAGLENNI-LASKLADADQNIQOQYAHQKKEKMLELELADLQVQHQE--KTAQEQ 298
Qy 339 HQREKDFLLKEAVESQRMCELMKQETHLKOALYTKFEPFNTLSKSSEVFTTFKQE 398
Db 299 TQWQ-----LYADQVSQLM-STEKNLRLQLAADGERFOQFDALTKSNEVFETKKE 349
Qy 399 MEKWTKKIKLEKFTMYRSSESKALLEMAEKTVRDKELHGLQVKTQRIKLCRAL 458
Db 350 MEKQWKLKIDLEKDNFEMKSKCNSDVALLIIEBELMKKQVDFKFNQKELSLCRSL 409
Qy 459 QTER 462
Db 410 QABR 413

RESULT 18
Q8BWJ7 PRELIMINARY; PRT; 241 AA.
ID Q8BWJ7
AC Q8BWJ7
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Muscle-derived protein MDP77 variant 1 (Fragment).
GN 2310001N14RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
R2 SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK052324; BAC34937.1; -.
DR MGD; MGI:1916756; 2310001N14Rik.
FT NON_TER 241 241


```
SQ SEQUENCE 241 AA; 27232 MW; 0B2DD1BD215630BB CRC64;
Query Match 17.0%; Score 459; DB 11; Length 241;
Best Local Similarity 44.1%; Pred. No. 8.1e-15;
Matches 112; Conservative 25; Mismatches 67; Indels 50; Gaps 6;

Qy 18 PSQAAPAVEAGPGSSQAPRPEGAQARTAQSGAL-----RVSELSRQLEDILSTYC 71
Db 16 PPGSSSLNONGPCKQDGERCSTSGQA-PEQEGSLHPEKGAHDVABELSRQLEDIISTY- 73
Qy 72 VDNQGGPGEBCAAGEPAEPEDAESKRTTYARNGEPEPTPVYGEKPSGDPNTEIRQ 131
Db 74 -----GSAASPRGKESTSETKEQPNTEAPONEDVDYEE-----TTBEI-- 112
Qy 132 SDEVDGRHRRPQ-----EKQARGLGKEITLLMOTLNTLSTPPEKLA 175
Db 113 -----DREPTAPEPAAAKEPVSNKBOEKKILKGLGKEANILLQNLKQAPPEKLD 167
Qy 176 LCKKYAELLERHSNOKMLQKQSLQVQKDLHGEKSKAVLARSKLESLESLCRELQRH 235
Db 168 LFKYITELDHRTQKLLKLLKQQAQTOREKQDQSEHNRAVLAESKLESLESLCRELQRH 227
Qy 236 NRSLEKEGVQARE 249
Db 228 NKTLEETLQARE 241

RESULT 19
Q9RXD7 PRELIMINARY; PRT; 404 AA.
AC Q9RXD7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (At5G50840).
GN AT5G50840.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosida II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RA "Arabidopsis ORF clones";
RT Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY081318; AAL91207.1;
DR EMBL; EF008857; AAP68296.1;
KW Hypothetical protein.
SQ SEQUENCE 404 AA; 46796 MW; 3C31C38A7B437DF6 CRC64;

Query Match 16.7%; Score 450; DB 10; Length 404;
Best Local Similarity 31.4%; Pred. No. 3.7e-14;
Matches 127; Conservative 89; Mismatches 152; Indels 36; Gaps 11;

Qy 82 DGAQGEPA-EPEDAESKRTY-----VARNGEPEPTPVYGEKPSGDPNTEIROSDVG 136
Db 21 DGAETPLNPSKQTEETTHETVALEKTEKPTFFVPLCETDGNEDDEVADLIQBSIKL- 79

SQ SEQUENCE 241 AA; 27232 MW; 0B2DD1BD215630BB CRC64;
Query Match 17.0%; Score 459; DB 11; Length 241;
Best Local Similarity 44.1%; Pred. No. 8.1e-15;
Matches 112; Conservative 25; Mismatches 67; Indels 50; Gaps 6;

Qy 18 PSQAAPAVEAGPGSSQAPRPEGAQARTAQSGAL-----RVSELSRQLEDILSTYC 71
Db 16 PPGSSSLNONGPCKQDGERCSTSGQA-PEQEGSLHPEKGAHDVABELSRQLEDIISTY- 73
Qy 72 VDNQGGPGEBCAAGEPAEPEDAESKRTTYARNGEPEPTPVYGEKPSGDPNTEIRQ 131
Db 74 -----GSAASPRGKESTSETKEQPNTEAPONEDVDYEE-----TTBEI-- 112
Qy 132 SDEVDGRHRRPQ-----EKQARGLGKEITLLMOTLNTLSTPPEKLA 175
Db 113 -----DREPTAPEPAAAKEPVSNKBOEKKILKGLGKEANILLQNLKQAPPEKLD 167
Qy 176 LCKKYAELLERHSNOKMLQKQSLQVQKDLHGEKSKAVLARSKLESLESLCRELQRH 235
Db 168 LFKYITELDHRTQKLLKLLKQQAQTOREKQDQSEHNRAVLAESKLESLESLCRELQRH 227
Qy 236 NRSLEKEGVQARE 249
Db 228 NKTLEETLQARE 241

RESULT 19
Q9RXD7 PRELIMINARY; PRT; 404 AA.
AC Q9RXD7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (At5G50840).
GN AT5G50840.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosida II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RA "Arabidopsis ORF clones";
RT Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY081318; AAL91207.1;
DR EMBL; EF008857; AAP68296.1;
KW Hypothetical protein.
SQ SEQUENCE 404 AA; 46796 MW; 3C31C38A7B437DF6 CRC64;

Query Match 16.7%; Score 450; DB 10; Length 404;
Best Local Similarity 31.4%; Pred. No. 3.7e-14;
Matches 127; Conservative 89; Mismatches 152; Indels 36; Gaps 11;

Qy 82 DGAQGEPA-EPEDAESKRTY-----VARNGEPEPTPVYGEKPSGDPNTEIROSDVG 136
Db 21 DGAETPLNPSKQTEETTHETVALEKTEKPTFFVPLCETDGNEDDEVADLIQBSIKL- 79
```


152 IQKEI--TLLMOTINTLTPB-EKLAALCKYAEILLEHRSOKOMKLLQKKSQVQEK 208
1 MGNFDEAALLKSLF--GVDDAKYOKLTKLAE--SEKQNAELKIKVLVDYDKVVKVXDL 56
209 DILRGHSHKAVL-----ARSLKESLCRELQRNRLKEEGVQVAREEEERKEVTSHFQV 263
57 TEKLERNNQILLRTEBAKSLKLELCRGLQKQANHTREEACAKMKKLEVERGLAVEQLKV 116
264 TLNDIOLQMEHNRNRSKLRQENMLAELRLKLIQYB-----LREHIDKVKP 312
117 TLKDIEKTMABGRSKSDSLAENDKSLSEKSFSGHYEEMKVIQQOIKKEKYEBEYK 176
313 HDLQOOLVDKIQQAQOEMLEKAEERHOREKDFLLKEAVESQRMCELMKQOETHLQOLA 372
177 TWDLEIKLITAKLESASQYKKS-----GNEKDELAKVLEETARVGGALKTEKALREQVQ 232
373 LYTEKFESEPNLTLSKSESVFTTFKQEMERQMTKKIKLEKBTMTYMSRWESSNKALL---- 428
233 EYSAKYSELTSCLSKSNEAFDFKDEISRVNKKQVKEKGLSYKCKSDANKKVLVTM 292
429 ----EVAEKEKTVRDEKLEGLQVKIORLEKLCRALQ 459
293 TNOEYAEKIATSDK-----KIQMLENLCRALR 319
RESULT 23
Q9BZA4 PRELIMINARY; PRT; 181 AA.
AC Q9BZA4
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE L:popolysaccaride-specific response 5-like protein (Hypothetical
protein DKP2p451G0616).
GN DKP2p451G0616.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda-Kawaguchi T., Skaletsky H., Minx P.J., Brown L.G., Rozen S.,
Wilson R.K., Waterston R.H., Page D.C.;
RT "The DNA sequence of the human Y chromosome";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ansgorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
Osanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF332225; AKI3477.1; -
DR EMBL; AL832583; CAD89940.1; -
DR Genew; HGNC:18577; CYorf15B.
KW Hypothetical protein.
SQ SEQUENCE 181 AA; 20858 MW; 4BA8B830933012A1 CRC64;
Query Match 13.7%; Score 369.5; DB 4; Length 181;
Best Local Similarity 44.8%; Pred. No. 1.1e-10;
Matches 78; Conservative 31; Mismatches 46; Indels 19; Gaps 3;
360 MKQOETHLQKALYATEKFEFQNTLSKSESVFTTFKQEMERQMTKKIKLEKBTMTYMSRW 419
1 MKQSEALKEQLFLYMDKFEFQNTVAKTNELPFAKQETKLTXXIKLEKEMVITYTK 60
420 WESSKALLEMAEKEKTVRDEKLEGLQVKIORLEKLCRALQOTERNLKNRY----- 469
61 WENNNTTLLQMAEKEKTRDKNYKVFQIKLERLEKLYALQIERNEUSEKLGILKGVSVX 120
470 ---QDLISAGQGSGLTDSGPER-----RPEPGQAQAPSPR-VTEAPCYPGAPT 514
121 VADVDLAVPYTHSCADLSDSNMLNTSSKRAPGVHLRADPKGMVNEVKYSKALST 174

01-MAR-2003 (TREMELrel. 23, Created)
01-MAR-2003 (TREMELrel. 23, Last sequence update)
01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=22354683; PubMed=12466851;
RA THE FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs";
RL Nature 420:563-573 (2002).
DR EMBL; AK078192; BAC37168.1; -
KW Hypothetical protein.
FT NON TER 150
SQ SEQUENCE 150 AA; 16490 MW; C444717503B42E49 CRC64;
Query Match 14.5%; Score 392.5; DB 11; Length 150;
Best Local Similarity 57.1%; Pred. No. 7.3e-12;
Matches 88; Conservative 9; Mismatches 36; Indels 21; Gaps 4;
2 SSGQPPAGPEGAQERPSQAFAVEAEGPGSSQAPRKPEGAQARTAGSGLRDV----- 55
5 NTPTRRAARGNGKQDRE-----PMDDPQQLGRKPKVPOARLL--GRPGLELKLQS 56
56 -----SSELROLEDILTSTYCVNNQGGPGEAGQEPAPSDAERSRYVARNGEPEP 109
57 LGPSVTSILRSRLQLEDILTSTYCVNNQGGPGEAGQEPTEPEDEKTSRYVARNGEPEP 116
110 -TPVYGEKESKGDPTNTEIRQSDVGDHRHR 142
117 GIPVNGEKETSGEPGTETIRASDEVGDRDHR 150
RESULT 22
Q22666 PRELIMINARY; PRT; 335 AA.
AC Q22666
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE T22C1.6 protein.
GN T22C1.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMuray A.A.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
investigating biology";
RL Science 282:2012-2018 (1998).
DR EMBL; Z75550; CAA99923.2; -
DR PIR; F87844; F87844.
DR PIR; T25110; T25110.
DR WormPep; T22C1.6; C2524002.
SQ SEQUENCE 335 AA; 38338 MW; DFC45011AD2FD2F5 CRC64;
Query Match 13.8%; Score 372.5; DB 5; Length 335;
Best Local Similarity 30.8%; Pred. No. 1.5e-10;
Matches 103; Conservative 70; Mismatches 120; Indels 41; Gaps 9;

```
RESULT 24
ID Q9FGD8 PRELIMINARY; PRT; 372 AA.
AC Q9FGD8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Genomic DNA, chromosome 5, TAC clone:K16E14.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurasids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RC Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026637; BAB10777.1;
DR EMBL; AB025617; BAB10777.1; JOINED.
SQ SEQUENCE 372 AA; 43086 MW; A706FP8CF3C75576 CRC64;

Query Match 12.3%; Score 332; DB 10; Length 372;
Best Local Similarity 26.8%; Pred. No. 1.4e-08;
Matches 106; Conservative 81; Mismatches 136; Indels 72; Gaps 13;

QY 109 PTPVYVGEKPSKGDPT-REIQSDVGDVDRHRRPOEKKAKGLGKEITLLMTLTL 167
DQ 16 PDGFDGATPEPLNSPKTQETTHETVAIEKTEPR-----TFP 55
QY 168 TPKEKLAALCK-----KYAELEHNSKQKMLQKQSLVQEKDHLRG-----B 214
DQ 56 VP-----LCETDGNEDDEADVADIQESIKLEFEQKEKSPISQYSDVSDICADDIE 109
QY 215 HSKAVLRSKLESICRELQRN-----RSLKRGVQRAREEERKEVTSHP 261
DQ 110 SRRLVRSVRVQRICSN-QRKSFWNLCSISKSLKEILBECKRVSTEGTILRSGLSTKP 168
QY 262 QVTLLNDIQLQM-EQHNERNKLRQNMELARLKLIEQYELREHIDKVFHKDLQOQL 320
DQ 169 QDAIMDVSIKLDQKNESLTQLK-ENEMLRTKLHLADQFMLSQEQHEQRLKQKLTLEQI 227
QY 321 VDALKQIQEMLKEAEERHOREKDFLKEAVESQRCMLKQOETHLQOQLALYTERKEE 380
DQ 228 SALKIKOHEKJL--THEOSQMK-----VYADQVSQLL-STENKRLQLTSDGDKFQQ 276
QY 381 FQNTLSKSEVPTFKQEMKMTKKIKLEKETTMYRSWRSSNKALLEAEKTVRDK 440
DQ 277 FQDALVKSNEVETFKQEDIDKMSKALKELRKENAFKNTKSDITLIELVEERBLKKL 336
QY 441 LEGLQVKIQRLLKLCALQTER-----NDLNKRVQ 470
DQ 337 LEKTKQKQKLSLCSLQAEKRRKQKRYTNSDTSVQ 371

RESULT 25
Q9VTUO PRELIMINARY; PRT; 1168 AA.
ID Q9VTUO;
AC Q9VTUO;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG32652 protein.
DE CG32652 OR CG1905.
GN Drosophila melanogaster (fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
```

```
QY 61 RTAQSALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQEPAPEDAEKSRITYVARN 120
DB 61 RTAQSALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQEPAPEDAEKSRITYVARN 120
QY 121 GEPEPTPVVNGEKPSPKDPNTEIEIRQSDVGDHRRPQKKAAGLKEITLLMQTLN 180
DB 121 GEPEPTPVVNGEKPSPKDPNTEIEIRQSDVGDHRRPQKKAAGLKEITLLMQTLN 180
QY 181 TLTSPPEKLAALCKKYAELLBEHNSQKMKLLQKQSQLVQEKDHLRGHSHKAVLARSK 240
DB 181 TLTSPPEKLAALCKKYAELLBEHNSQKMKLLQKQSQLVQEKDHLRGHSHKAVLARSK 240
QY 241 LESLCRELQNRNRLSKBEGVQARABEEERKEVTSHFQVTLNDIQLQMEQHNEHNSKLRQ 300
DB 241 LESLCRELQNRNRLSKBEGVQARABEEERKEVTSHFQVTLNDIQLQMEQHNEHNSKLRQ 300
QY 301 ENMELAEKRLKGLIEQYELREBEHIDKVFHKDLQQLQVDAKLQQAQEMLEAEERHOREKD 360
DB 301 ENMELAEKRLKGLIEQYELREBEHIDKVFHKDLQQLQVDAKLQQAQEMLEAEERHOREKD 360
QY 361 FLLKEAVESQRMCELMKQOETHLQKQALYTEKFEFQNTLSKSESVFTTFKQEMEKMTK 420
DB 361 FLLKEAVESQRMCELMKQOETHLQKQALYTEKFEFQNTLSKSESVFTTFKQEMEKMTK 420
QY 421 KIKLEKETTTMYRSRWESSNKALLEMAEKTVRDKLEGLQVQKIQLEKLCRALQTERND 480
DB 421 KIKLEKETTTMYRSRWESSNKALLEMAEKTVRDKLEGLQVQKIQLEKLCRALQTERND 480
QY 481 LNKRVQDLSAGQGSGLTDSGPERPPEGGAQAPSPRVTEAPCPYGPAPSTEASQGTGPQE 540
DB 481 LNKRVQDLSAGQGSGLTDSGPERPPEGGAQAPSPRVTEAPCPYGPAPSTEASQGTGPQE 540
QY 541 PTSARA 546
DB 541 PTSARA 546
```

RESULT 2

```
US-10-671-242-44
; Sequence 44, Application US/10671242
; Publication No. US2004004049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; PRIOR FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-44
```

Query Match 100.0%; Score 546; DB 12; Length 546;
Best Local Similarity 100.0%; Pred. No. 0;

```
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKNQDKNGAAGKQSNFKSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKPEGAQA 60
DB 1 MKNQDKNGAAGKQSNFKSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKPEGAQA 60
QY 61 RTAQSALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQEPAPEDAEKSRITYVARN 120
DB 61 RTAQSALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQEPAPEDAEKSRITYVARN 120
QY 121 GEPEPTPVVNGEKPSPKDPNTEIEIRQSDVGDHRRPQKKAAGLKEITLLMQTLN 180
DB 121 GEPEPTPVVNGEKPSPKDPNTEIEIRQSDVGDHRRPQKKAAGLKEITLLMQTLN 180
QY 181 TLTSPPEKLAALCKKYAELLBEHNSQKMKLLQKQSQLVQEKDHLRGHSHKAVLARSK 240
DB 181 TLTSPPEKLAALCKKYAELLBEHNSQKMKLLQKQSQLVQEKDHLRGHSHKAVLARSK 240
QY 241 LESLCRELQNRNRLSKBEGVQARABEEERKEVTSHFQVTLNDIQLQMEQHNEHNSKLRQ 300
DB 241 LESLCRELQNRNRLSKBEGVQARABEEERKEVTSHFQVTLNDIQLQMEQHNEHNSKLRQ 300
QY 301 ENMELAEKRLKGLIEQYELREBEHIDKVFHKDLQQLQVDAKLQQAQEMLEAEERHOREKD 360
DB 301 ENMELAEKRLKGLIEQYELREBEHIDKVFHKDLQQLQVDAKLQQAQEMLEAEERHOREKD 360
QY 361 FLLKEAVESQRMCELMKQOETHLQKQALYTEKFEFQNTLSKSESVFTTFKQEMEKMTK 420
DB 361 FLLKEAVESQRMCELMKQOETHLQKQALYTEKFEFQNTLSKSESVFTTFKQEMEKMTK 420
QY 421 KIKLEKETTTMYRSRWESSNKALLEMAEKTVRDKLEGLQVQKIQLEKLCRALQTERND 480
DB 421 KIKLEKETTTMYRSRWESSNKALLEMAEKTVRDKLEGLQVQKIQLEKLCRALQTERND 480
QY 481 LNKRVQDLSAGQGSGLTDSGPERPPEGGAQAPSPRVTEAPCPYGPAPSTEASQGTGPQE 540
DB 481 LNKRVQDLSAGQGSGLTDSGPERPPEGGAQAPSPRVTEAPCPYGPAPSTEASQGTGPQE 540
QY 541 PTSARA 546
DB 541 PTSARA 546
```

RESULT 3

```
US-10-023-529-44
; Sequence 44, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
```

; ORGANISM: Homo sapiens
US-10-023-529-44

Query Match 100.0%; Score 546; DB 13; Length 546;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKQDKKGAAGKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPKPEGAQA 60
DB 1 MKQDKKGAAGKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPKPEGAQA 60

QY 61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGGPGEDGAQEPAPDAEKSRRTYVARN 120
DB 61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGGPGEDGAQEPAPDAEKSRRTYVARN 120

QY 121 GEPEPTPVNGEKEPSKGDPTNTEIRQSDVGVDRDHRPQEKKAAGLKGKETTLLMOTLN 180
DB 121 GEPEPTPVNGEKEPSKGDPTNTEIRQSDVGVDRDHRPQEKKAAGLKGKETTLLMOTLN 180

QY 181 TLSTPEKLAALCKKYAELLBEHRNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSK 240
DB 181 TLSTPEKLAALCKKYAELLBEHRNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSK 240

QY 241 LESLCHRELQHNRSLSKEGVQVAREEKEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
DB 241 LESLCHRELQHNRSLSKEGVQVAREEKEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300

QY 301 ENMELAEKLLIQVELREEHIDKVFKHDLQQLVDAKLQQAQEMLKEAEERHOREKD 360
DB 301 ENMELAEKLLIQVELREEHIDKVFKHDLQQLVDAKLQQAQEMLKEAEERHOREKD 360

QY 361 FLKKEAVESQRMCELMKQOETHLQKQALYTEKPEFQNTLSKSSSEVFTTFKQEMKMTK 420
DB 361 FLKKEAVESQRMCELMKQOETHLQKQALYTEKPEFQNTLSKSSSEVFTTFKQEMKMTK 420

QY 421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERND 480
DB 421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERND 480

QY 481 LNKRVQDLSAGGQSLTDSGPERPPEGPAQAPSSPRVTEAPCYGAPSTASGQTGPQE 540
DB 481 LNKRVQDLSAGGQSLTDSGPERPPEGPAQAPSSPRVTEAPCYGAPSTASGQTGPQE 540

QY 541 PTSARA 546
DB 541 PTSARA 546

RESULT 5
US-10-616-187-44
; Sequence 44, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608

; ORGANISM: Homo sapiens
US-10-023-529-44

Query Match 100.0%; Score 546; DB 13; Length 546;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKQDKKGAAGKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPKPEGAQA 60
DB 1 MKQDKKGAAGKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPKPEGAQA 60

QY 61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGGPGEDGAQEPAPDAEKSRRTYVARN 120
DB 61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGGPGEDGAQEPAPDAEKSRRTYVARN 120

QY 121 GEPEPTPVNGEKEPSKGDPTNTEIRQSDVGVDRDHRPQEKKAAGLKGKETTLLMOTLN 180
DB 121 GEPEPTPVNGEKEPSKGDPTNTEIRQSDVGVDRDHRPQEKKAAGLKGKETTLLMOTLN 180

QY 181 TLSTPEKLAALCKKYAELLBEHRNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSK 240
DB 181 TLSTPEKLAALCKKYAELLBEHRNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSK 240

QY 241 LESLCHRELQHNRSLSKEGVQVAREEKEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
DB 241 LESLCHRELQHNRSLSKEGVQVAREEKEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300

QY 301 ENMELAEKLLIQVELREEHIDKVFKHDLQQLVDAKLQQAQEMLKEAEERHOREKD 360
DB 301 ENMELAEKLLIQVELREEHIDKVFKHDLQQLVDAKLQQAQEMLKEAEERHOREKD 360

QY 361 FLKKEAVESQRMCELMKQOETHLQKQALYTEKPEFQNTLSKSSSEVFTTFKQEMKMTK 420
DB 361 FLKKEAVESQRMCELMKQOETHLQKQALYTEKPEFQNTLSKSSSEVFTTFKQEMKMTK 420

QY 421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERND 480
DB 421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERND 480

QY 481 LNKRVQDLSAGGQSLTDSGPERPPEGPAQAPSSPRVTEAPCYGAPSTASGQTGPQE 540
DB 481 LNKRVQDLSAGGQSLTDSGPERPPEGPAQAPSSPRVTEAPCYGAPSTASGQTGPQE 540

QY 541 PTSARA 546
DB 541 PTSARA 546

RESULT 4
US-10-023-523-44
; Sequence 44, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03

NUMBER OF SEQ ID NOS: 2700
SOFTWARE: Custom
SEQ ID NO 2134
LENGTH: 510
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(510)
OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-2134

Query Match 87.2%; Score 476; DB 12; Length 510;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNQDKKNGAAGKQSPKSSPGQPEAGPEGAQERPSQAAPAVAEAGPGSSQAAPKPEGAQA 60
DB 12 MNQDKKNGAAGKQSPKSSPGQPEAGPEGAQERPSQAAPAVAEAGPGSSQAAPKPEGAQA 71
QY 61 RTAQSGALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYARN 120
DB 72 RTAQSGALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYARN 131
QY 121 GSEPTPVNNGEKPSPKSGDPNTEIRQSDVGDHRRRPOEKKAKGLGKEITLLMQTLN 180
DB 132 GSEPTPVNNGEKPSPKSGDPNTEIRQSDVGDHRRRPOEKKAKGLGKEITLLMQTLN 191
QY 181 TLSTPEEKLAALCKKYAELLBEHRNSQKMLLQKQSQVLQVQEKDHLRGEHSAVLARSK 240
DB 192 TLSTPEEKLAALCKKYAELLBEHRNSQKMLLQKQSQVLQVQEKDHLRGEHSAVLARSK 251
QY 241 LESLCRLQRNRLSKBEGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
DB 252 LESLCRLQRNRLSKBEGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQ 311
QY 301 ENMELAEKLLIIOYELREEHIDKVFHKDLQQLQVDAKLQQAQEMLEABERHOREKD 360
DB 312 ENMELAEKLLIIOYELREEHIDKVFHKDLQQLQVDAKLQQAQEMLEABERHOREKD 371
QY 361 PLLKEAVESQRMCELMKQOETHLKOALALYTKPEFQNTLSKSEVFTTFKQEMEKMTK 420
DB 372 PLLKEAVESQRMCELMKQOETHLKOALALYTKPEFQNTLSKSEVFTTFKQEMEKMTK 431
QY 421 KIKLEKETTYRWRWESSNKALLEMAEKTVDKLEGLQVKTORLEKLCRALQ 476
DB 432 KIKLEKETTYRWRWESSNKALLEMAEKTVDKLEGLQVKTORLEKLCRALQ 487

RESULT 7

US-09-962-055-8

Sequence 8, Application US/09962055

Patent No. US30020052033A1

GENERAL INFORMATION:

APPLICANT: Lees, Ann M. S.

Lees, Robert S.

Law, Simon W.

Arjona, Anibal A.

TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN

BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND

TREATING ATHEROSCLEROSIS

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 44
LENGTH: 546
TYPE: PRT
ORGANISM: Homo sapiens
US-10-616-187-44

Query Match 100.0%; Score 546; DB 15; Length 546;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNQDKKNGAAGKQSPKSSPGQPEAGPEGAQERPSQAAPAVAEAGPGSSQAAPKPEGAQA 60
DB 1 MNQDKKNGAAGKQSPKSSPGQPEAGPEGAQERPSQAAPAVAEAGPGSSQAAPKPEGAQA 60
QY 61 RTAQSGALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYARN 120
DB 61 RTAQSGALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYARN 120
QY 121 GSEPTPVNNGEKPSPKSGDPNTEIRQSDVGDHRRRPOEKKAKGLGKEITLLMQTLN 180
DB 121 GSEPTPVNNGEKPSPKSGDPNTEIRQSDVGDHRRRPOEKKAKGLGKEITLLMQTLN 180
QY 181 TLSTPEEKLAALCKKYAELLBEHRNSQKMLLQKQSQVLQVQEKDHLRGEHSAVLARSK 240
DB 181 TLSTPEEKLAALCKKYAELLBEHRNSQKMLLQKQSQVLQVQEKDHLRGEHSAVLARSK 240
QY 241 LESLCRLQRNRLSKBEGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
DB 241 LESLCRLQRNRLSKBEGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
QY 301 ENMELAEKLLIIOYELREEHIDKVFHKDLQQLQVDAKLQQAQEMLEABERHOREKD 360
DB 301 ENMELAEKLLIIOYELREEHIDKVFHKDLQQLQVDAKLQQAQEMLEABERHOREKD 360
QY 361 PLLKEAVESQRMCELMKQOETHLKOALALYTKPEFQNTLSKSEVFTTFKQEMEKMTK 420
DB 361 PLLKEAVESQRMCELMKQOETHLKOALALYTKPEFQNTLSKSEVFTTFKQEMEKMTK 420
QY 421 KIKLEKETTYRWRWESSNKALLEMAEKTVDKLEGLQVKTORLEKLCRALQ 480
DB 421 KIKLEKETTYRWRWESSNKALLEMAEKTVDKLEGLQVKTORLEKLCRALQ 480
QY 481 LNKRVQDLSAGQGLTDSGPERPEGFGAQPSSPRVTPAPCPYPGAPSTEASQGTGPQE 540
DB 481 LNKRVQDLSAGQGLTDSGPERPEGFGAQPSSPRVTPAPCPYPGAPSTEASQGTGPQE 540
QY 541 PTSARA 546
DB 541 PTSARA 546

RESULT 6

US-10-276-774-2134

Sequence 2134, Application US/10276774

Publication No. US20040053245A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

APPLICANT: Tang, Y, Tom et al

TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides

FILE REFERENCE: 21272-030

CURRENT APPLICATION NUMBER: US/10/276,774

CURRENT FILING DATE: 2002-11-18

PRIOR APPLICATION NUMBER: 09/560,875

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: 09/496,914

PRIOR FILING DATE: 2000-02-03

SOFTWARE: FastSeq for Windows Version 2.0

APPLICANT: Law, Simon W.
APPLICANT: Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
TITLE OF INVENTION: ATHEROSCLEROSIS
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/09/976,740
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 09/616,289
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 08/979,608
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 530
TYPE: PRT
ORGANISM: Homo sapiens
US-09-976-740-8

Query Match 78.6%; Score 429; DB 9; Length 530;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 17 KSPGQPEAGPEGAQRPQAAPVPAEAGPGSSQAPKPEGAQARTAGSALRDVSEELS 76
DB 1 KSPGQPEAGPEGAQRPQAAPVPAEAGPGSSQAPKPEGAQARTAGSALRDVSEELS 60
QY 77 RQLEDILSTYCVNNQGGGEGDGAQGPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 136
DB 61 RQLEDILSTYCVNNQGGGEGDGAQGPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 120
QY 137 KGPNTTEIRQSDVDRDRHRRPQEKKAGLGEITLLMQTLNTLSTPEEKLAALCKKY 196
DB 121 KGPNTTEIRQSDVDRDRHRRPQEKKAGLGEITLLMQTLNTLSTPEEKLAALCKKY 180
QY 197 AELLEHRNSQKMLQKQSQVLQVQKHDLRGHSHKAVLARSKLESICREIQNRSLK 256
DB 181 AELLEHRNSQKMLQKQSQVLQVQKHDLRGHSHKAVLARSKLESICREIQNRSLK 240
QY 257 EGVQVQAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIQY 316
DB 241 EGVQVQAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIQY 300
QY 317 ELREEHIDKVPFKHDLQQLVDKQLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376
DB 301 ELREEHIDKVPFKHDLQQLVDKQLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
QY 377 KQOETHLQKQALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 436
DB 361 KQOETHLQKQALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
QY 437 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSLS 496
DB 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSLS 480
QY 497 TDSGPRRPEGGAQAPSSPRVTEAPCPGAPSTASGQTGPQPTTSARA 546
DB 481 TDSGPRRPEGGAQAPSSPRVTEAPCPGAPSTASGQTGPQPTTSARA 530

RESULT 8

US-09-976-740-8
Sequence 8, Application US/09976740
Publication No. US20020194633A1
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
APPLICANT: Lees, Robert S.
APPLICANT: Law, Simon W.

Query Match 78.6%; Score 429; DB 9; Length 530;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 17 KSPGQPEAGPEGAQRPQAAPVPAEAGPGSSQAPKPEGAQARTAGSALRDVSEELS 76
DB 1 KSPGQPEAGPEGAQRPQAAPVPAEAGPGSSQAPKPEGAQARTAGSALRDVSEELS 60
QY 77 RQLEDILSTYCVNNQGGGEGDGAQGPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 136
DB 61 RQLEDILSTYCVNNQGGGEGDGAQGPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 120
QY 137 KGPNTTEIRQSDVDRDRHRRPQEKKAGLGEITLLMQTLNTLSTPEEKLAALCKKY 196
DB 121 KGPNTTEIRQSDVDRDRHRRPQEKKAGLGEITLLMQTLNTLSTPEEKLAALCKKY 180
QY 197 AELLEHRNSQKMLQKQSQVLQVQKHDLRGHSHKAVLARSKLESICREIQNRSLK 256
DB 181 AELLEHRNSQKMLQKQSQVLQVQKHDLRGHSHKAVLARSKLESICREIQNRSLK 240
QY 257 EGVQVQAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIQY 316
DB 241 EGVQVQAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIQY 300
QY 317 ELREEHIDKVPFKHDLQQLVDKQLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376
DB 301 ELREEHIDKVPFKHDLQQLVDKQLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
QY 377 KQOETHLQKQALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 436
DB 361 KQOETHLQKQALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
QY 437 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSLS 496
DB 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSLS 480
QY 497 TDSGPRRPEGGAQAPSSPRVTEAPCPGAPSTASGQTGPQPTTSARA 546
DB 481 TDSGPRRPEGGAQAPSSPRVTEAPCPGAPSTASGQTGPQPTTSARA 530

RESULT 9

US-10-671-242-8
Sequence 8, Application US/10671242
Publication No. US20040040049A1
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
APPLICANT: Lees, Robert S.
APPLICANT: Law, Simon W.

RESULT 11
US-10-023-523-8
; Sequence 8, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.

; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-523-8

Query Match 78.6%; Score 429; DB 13; Length 530;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 17 KSSPGQPEAGPEGAQRRPSQAAPAVAEAGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 76
Db 1 KSSPGQPEAGPEGAQRRPSQAAPAVAEAGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 60
QY 77 RQLEDILSTYCVDNNGGPGEDGAGQEPAPEDAEKSTTYVANGPEPEPTPVYNGEKPS 136
Db 61 RQLEDILSTYCVDNNGGPGEDGAGQEPAPEDAEKSTTYVANGPEPEPTPVYNGEKPS 120
QY 137 KGPNTTEIRQSDVGDHRRPOEKKKAGLGKEITLLMQTLNTLSTPEEKLAALCKKY 196
Db 121 KGPNTTEIRQSDVGDHRRPOEKKKAGLGKEITLLMQTLNTLSTPEEKLAALCKKY 180
QY 197 AELLLEHNSQKMKLLQKQSQLVQEKDHLRGHNSKAVLARSKLSLCSRLQHNRSJK 256
Db 181 AELLLEHNSQKMKLLQKQSQLVQEKDHLRGHNSKAVLARSKLSLCSRLQHNRSJK 240
QY 257 EGVQARABEEERKEVTSHFQVTLNDIQLQMEQHNSKLRQENNELAERLKKLIEQY 316
Db 241 EGVQARABEEERKEVTSHFQVTLNDIQLQMEQHNSKLRQENNELAERLKKLIEQY 300
QY 317 ELREEHIDKVFHKDLQQLVDKLAQQAQEMLKEAERHOREKDPLLKEAVESQRMCELM 376
Db 301 ELREEHIDKVFHKDLQQLVDKLAQQAQEMLKEAERHOREKDPLLKEAVESQRMCELM 360
QY 377 KQETHLKOQALALYTEKEEFQNTLSKSEVFTTTPKQEMKMTKKIKKLEKTTMYRSW 436
Db 361 KQETHLKOQALALYTEKEEFQNTLSKSEVFTTTPKQEMKMTKKIKKLEKTTMYRSW 420
QY 437 ESSNKALLEMAEEKTVRDKELEGLOVKIORLEKLCRALQTERNDLNKRVDLSAGQGSGL 496
Db 421 ESSNKALLEMAEEKTVRDKELEGLOVKIORLEKLCRALQTERNDLNKRVDLSAGQGSGL 480
QY 497 TDGSPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 546
Db 481 TDGSPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 530

RESULT 12
US-10-616-187-8
; Sequence 8, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:

; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-8

Query Match 78.6%; Score 429; DB 15; Length 530;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 17 KSSPGQPEAGPEGAQRRPSQAAPAVAEAGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 76
Db 1 KSSPGQPEAGPEGAQRRPSQAAPAVAEAGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 60
QY 77 RQLEDILSTYCVDNNGGPGEDGAGQEPAPEDAEKSTTYVANGPEPEPTPVYNGEKPS 136
Db 61 RQLEDILSTYCVDNNGGPGEDGAGQEPAPEDAEKSTTYVANGPEPEPTPVYNGEKPS 120
QY 137 KGPNTTEIRQSDVGDHRRPOEKKKAGLGKEITLLMQTLNTLSTPEEKLAALCKKY 196
Db 121 KGPNTTEIRQSDVGDHRRPOEKKKAGLGKEITLLMQTLNTLSTPEEKLAALCKKY 180
QY 197 AELLLEHNSQKMKLLQKQSQLVQEKDHLRGHNSKAVLARSKLSLCSRLQHNRSJK 256
Db 181 AELLLEHNSQKMKLLQKQSQLVQEKDHLRGHNSKAVLARSKLSLCSRLQHNRSJK 240
QY 257 EGVQARABEEERKEVTSHFQVTLNDIQLQMEQHNSKLRQENNELAERLKKLIEQY 316
Db 241 EGVQARABEEERKEVTSHFQVTLNDIQLQMEQHNSKLRQENNELAERLKKLIEQY 300
QY 317 ELREEHIDKVFHKDLQQLVDKLAQQAQEMLKEAERHOREKDPLLKEAVESQRMCELM 376
Db 301 ELREEHIDKVFHKDLQQLVDKLAQQAQEMLKEAERHOREKDPLLKEAVESQRMCELM 360
QY 377 KQETHLKOQALALYTEKEEFQNTLSKSEVFTTTPKQEMKMTKKIKKLEKTTMYRSW 436
Db 361 KQETHLKOQALALYTEKEEFQNTLSKSEVFTTTPKQEMKMTKKIKKLEKTTMYRSW 420
QY 437 ESSNKALLEMAEEKTVRDKELEGLOVKIORLEKLCRALQTERNDLNKRVDLSAGQGSGL 496
Db 421 ESSNKALLEMAEEKTVRDKELEGLOVKIORLEKLCRALQTERNDLNKRVDLSAGQGSGL 480
QY 497 TDGSPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 546
Db 481 TDGSPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 530

RESULT 13
US-09-962-055-5
; Sequence 5, Application US/09962055
; Patent No. US200205203A1

```
;
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
;
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,055
; FILING DATE: 24-Sep-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-962-055-5
Query Match 31.5%; Score 172; DB 9; Length 557;
Best Local Similarity 100.0%; Pred. No. 6.7e-156;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 280 TLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQYELREHIDKVFKEHDLQOQLVDA 339
DB 280 TLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQYELREHIDKVFKEHDLQOQLVDA 339
QY 340 KLOQAQEMLKAEERHOREKDFLLKEAVESQRMCELMKQOETHLKOOLALYTEKFEFQN 399
DB 340 KLOQAQEMLKAEERHOREKDFLLKEAVESQRMCELMKQOETHLKOOLALYTEKFEFQN 399
QY 400 TLSKSSVFTTFKQEMERKTKIKKLEKTTMYRSWESSNKALLEMAEKT 451
DB 400 TLSKSSVFTTFKQEMERKTKIKKLEKTTMYRSWESSNKALLEMAEKT 451
RESULT 15
US-10-671-242-5
; Sequence 5, Application US/10671242
; Publication No. US20040040049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; CURRENT FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-671-242-5
Query Match 31.5%; Score 172; DB 12; Length 557;
Best Local Similarity 100.0%; Pred. No. 6.7e-156;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 280 TLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQYELREHIDKVFKEHDLQOQLVDA 339
DB 280 TLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQYELREHIDKVFKEHDLQOQLVDA 339
QY 340 KLOQAQEMLKAEERHOREKDFLLKEAVESQRMCELMKQOETHLKOOLALYTEKFEFQN 399
DB 340 KLOQAQEMLKAEERHOREKDFLLKEAVESQRMCELMKQOETHLKOOLALYTEKFEFQN 399
QY 400 TLSKSSVFTTFKQEMERKTKIKKLEKTTMYRSWESSNKALLEMAEKT 451
DB 400 TLSKSSVFTTFKQEMERKTKIKKLEKTTMYRSWESSNKALLEMAEKT 451
RESULT 14
US-09-976-740-5
; Sequence 5, Application US/09976740
; Publication No. US2002019463A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
```

	Matches	172;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	280	TLNDIQLQVQEQHNRNS	KLQENMELAEKLI	EOYELREEHIDKVF	KHKLQQLVDA	339				
Db	280	TLNDIQLQVQEQHNRNS	KLQENMELAEKLI	EOYELREEHIDKVF	KHKLQQLVDA	339				
Qy	340	KLQQAQEMLKAEERHORE	KDFLLKEAVESQRMCE	LMKQOETHLXQQLALY	TEKFEFQN	399				
Db	340	KLQQAQEMLKAEERHORE	KDFLLKEAVESQRMCE	LMKQOETHLXQQLALY	TEKFEFQN	399				
Qy	400	TLKSSEVFTTQEMEKMT	KKIKLEKETTMYRSRW	SSNKALLEMEBEKT	451					
Db	400	TLKSSEVFTTQEMEKMT	KKIKLEKETTMYRSRW	SSNKALLEMEBEKT	451					

Search completed: June 8, 2004, 16:41:32
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:41:31 ; Search time 40 Seconds
(without alignments)
4180.616 Million cell updates/sec

Title: US-10-023-529-8
Perfect score: 530
Sequence: 1 KSSPQPRAGPAGPQRPSPQ.....APSTEASQTPQRPPTSARA 530

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 300 summaries

Database : SPTRMEL.25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	429	80.9	546	4 Q86VW3	Q86VW3 homo sapien
2	412	77.7	546	4 Q86T54	Q86T54 homo sapien
3	328	61.9	546	4 Q86T86	Q86T86 homo sapien
4	178	33.6	546	4 Q86T85	Q86T85 homo sapien
5	116	21.9	116	4 Q86Y86	Q86Y86 homo sapien
6	21	4.0	150	11 Q8C5K1	Q8C5K1 mus musculus
7	19	3.6	241	11 Q8BJW7	Q8BJW7 mus musculus
8	19	3.6	241	11 Q8BP11	Q8BP11 mus musculus
9	19	3.6	463	11 Q8BUK2	Q8BUK2 mus musculus
10	19	3.6	524	11 Q8BHN1	Q8BHN1 mus musculus
11	19	3.6	528	4 Q8NUQ3	Q8NUQ3 homo sapien
12	19	3.6	684	11 Q8VBT1	Q8VBT1 mus musculus
13	17	3.2	186	4 Q8POX1	Q8POX1 homo sapien
14	17	3.2	505	4 Q8X3S2	Q8X3S2 homo sapien
15	17	3.2	676	13 Q8I969	Q8I969 gallus gall
16	17	3.2	715	4 Q8N3L3	Q8N3L3 homo sapien

90	7	1.3	74	12	Q9W1K7	Q9w1k7 hepatitis c	163	7	1.3	172	17	O28639	O28639 archaeoglob
91	7	1.3	75	12	Q9W916	Q9w916 hepatitis c	164	7	1.3	173	4	Q9Y5A0	Q9y5a0 homo sapien
92	7	1.3	76	12	Q9Z8F1	Q9z8f1 listeria in	165	7	1.3	176	5	Q9Y5A0	Q9y5a0 schistosoma
93	7	1.3	77	12	Q9Z982	Q9z982 hepatitis c	166	7	1.3	177	5	O76627	O76627 caenorhabdi
94	7	1.3	78	12	Q9RSW4	Q9rsW4 deinococcus	167	7	1.3	177	16	O8G322	O8g322 bifidobacte
95	7	1.3	79	16	Q9S3L1	Q9s3l1 shigella fl	168	7	1.3	179	10	O23356	O23356 arabidopsis
96	7	1.3	80	16	Q8W81	O8w81 pseudomonas	169	7	1.3	180	16	O8Y7H7	O8y7h7 brucella me
97	7	1.3	81	2	O66133	O66133 acidiphiliu	170	7	1.3	181	16	O50914	O50914 borrelia bu
98	7	1.3	82	2	O66133	O66133 acidiphiliu	171	7	1.3	181	16	O50914	O50914 borrelia bu
99	7	1.3	83	16	O92J77	O92j77 rhizobium m	172	7	1.3	186	2	Q9AEN4	Q9aen4 burkholderi
100	7	1.3	84	16	O92J77	O92j77 rhizobium m	173	7	1.3	187	2	Q9AEN4	Q9aen4 burkholderi
101	7	1.3	85	2	O66135	O66135 acidiphiliu	174	7	1.3	187	16	Q92XC5	Q92xc5 rhizobium m
102	7	1.3	86	16	O66135	O66135 acidiphiliu	175	7	1.3	190	1	P94913	P94913 methanosarc
103	7	1.3	87	10	O8L5Q4	O8l5q4 cicier ariet	176	7	1.3	190	1	P94913	P94913 methanosarc
104	7	1.3	88	16	O8WR28	O8wr28 anopheles g	177	7	1.3	193	17	Q96XQ6	Q96xq6 sulfolobus
105	7	1.3	89	16	O8WR28	O8wr28 anopheles g	178	7	1.3	193	17	Q96XQ6	Q96xq6 sulfolobus
106	7	1.3	90	16	O8WR28	O8wr28 anopheles g	179	7	1.3	193	17	Q96XQ6	Q96xq6 sulfolobus
107	7	1.3	91	16	O8WR28	O8wr28 anopheles g	180	7	1.3	193	17	Q96XQ6	Q96xq6 sulfolobus
108	7	1.3	92	16	O8WR28	O8wr28 anopheles g	181	7	1.3	193	17	Q96XQ6	Q96xq6 sulfolobus
109	7	1.3	93	16	O8WR28	O8wr28 anopheles g	182	7	1.3	193	17	Q96XQ6	Q96xq6 sulfolobus
110	7	1.3	94	3	P79038	P79038 emericeila	183	7	1.3	194	12	O91DH7	O91dh7 hepatitis d
111	7	1.3	95	3	P79038	P79038 emericeila	184	7	1.3	194	12	O91DH7	O91dh7 hepatitis d
112	7	1.3	96	15	Q9Q4Y7	Q9q4y7 human immun	185	7	1.3	194	12	O91DH7	O91dh7 hepatitis d
113	7	1.3	97	10	O8H5T1	O8h5t1 oryza sativ	186	7	1.3	194	12	O91DH7	O91dh7 hepatitis d
114	7	1.3	98	10	O8H5T1	O8h5t1 oryza sativ	187	7	1.3	194	12	O91DH7	O91dh7 hepatitis d
115	7	1.3	99	10	O8H5T1	O8h5t1 oryza sativ	188	7	1.3	194	12	O91DH7	O91dh7 hepatitis d
116	7	1.3	100	12	O38021	O38021 potato viru	189	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
117	7	1.3	101	12	O38021	O38021 potato viru	190	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
118	7	1.3	102	12	O38021	O38021 potato viru	191	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
119	7	1.3	103	12	O38021	O38021 potato viru	192	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
120	7	1.3	104	12	O38021	O38021 potato viru	193	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
121	7	1.3	105	12	O38021	O38021 potato viru	194	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
122	7	1.3	106	12	O38021	O38021 potato viru	195	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
123	7	1.3	107	12	O38021	O38021 potato viru	196	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
124	7	1.3	108	12	O38021	O38021 potato viru	197	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
125	7	1.3	109	12	O38021	O38021 potato viru	198	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
126	7	1.3	110	12	O38021	O38021 potato viru	199	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
127	7	1.3	111	12	O38021	O38021 potato viru	200	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
128	7	1.3	112	12	O38021	O38021 potato viru	201	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
129	7	1.3	113	12	O38021	O38021 potato viru	202	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
130	7	1.3	114	12	O38021	O38021 potato viru	203	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
131	7	1.3	115	12	O38021	O38021 potato viru	204	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
132	7	1.3	116	12	O38021	O38021 potato viru	205	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
133	7	1.3	117	12	O38021	O38021 potato viru	206	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
134	7	1.3	118	12	O38021	O38021 potato viru	207	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
135	7	1.3	119	12	O38021	O38021 potato viru	208	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
136	7	1.3	120	12	O38021	O38021 potato viru	209	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
137	7	1.3	121	12	O38021	O38021 potato viru	210	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
138	7	1.3	122	12	O38021	O38021 potato viru	211	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
139	7	1.3	123	12	O38021	O38021 potato viru	212	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
140	7	1.3	124	12	O38021	O38021 potato viru	213	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
141	7	1.3	125	12	O38021	O38021 potato viru	214	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
142	7	1.3	126	12	O38021	O38021 potato viru	215	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
143	7	1.3	127	12	O38021	O38021 potato viru	216	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
144	7	1.3	128	12	O38021	O38021 potato viru	217	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
145	7	1.3	129	12	O38021	O38021 potato viru	218	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
146	7	1.3	130	12	O38021	O38021 potato viru	219	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
147	7	1.3	131	12	O38021	O38021 potato viru	220	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
148	7	1.3	132	12	O38021	O38021 potato viru	221	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
149	7	1.3	133	12	O38021	O38021 potato viru	222	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
150	7	1.3	134	12	O38021	O38021 potato viru	223	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
151	7	1.3	135	12	O38021	O38021 potato viru	224	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
152	7	1.3	136	12	O38021	O38021 potato viru	225	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
153	7	1.3	137	12	O38021	O38021 potato viru	226	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
154	7	1.3	138	12	O38021	O38021 potato viru	227	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
155	7	1.3	139	12	O38021	O38021 potato viru	228	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
156	7	1.3	140	12	O38021	O38021 potato viru	229	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
157	7	1.3	141	12	O38021	O38021 potato viru	230	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
158	7	1.3	142	12	O38021	O38021 potato viru	231	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
159	7	1.3	143	12	O38021	O38021 potato viru	232	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
160	7	1.3	144	12	O38021	O38021 potato viru	233	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
161	7	1.3	145	12	O38021	O38021 potato viru	234	7	1.3	195	12	Q9E326	Q9e326 hepatitis d
162	7	1.3	146	12	O38021	O38021 potato viru	235	7	1.3	195	12	Q9E326	Q9e326 hepatitis d


```
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
RA Koehler K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL; AL832338; CAD91138.1; -.
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61861 MW; 728067478997DF6 CRC64;

Query Match
Best Local Similarity 77.7%; Score 412; DB 4; Length 546;
Matches 512; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKPGEQAQTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKPGEQAQTAQSGALRDVSEELS 76
QY 61 RQLEDILSTYCVNNQGGPGEQAQERPSQAAPAVEAEGPGSSQAAPRKPGEQAQTAQSGALRDVSEELS 120
Db 77 RQLEDILSTYCVNNQGGPGEQAQERPSQAAPAVEAEGPGSSQAAPRKPGEQAQTAQSGALRDVSEELS 136
QY 121 KGDPTNTEIRQSDVEGDRHRRPQSKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 180
Db 137 KGDPTNTEIRQSDVEGDRHRRPQSKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 ABLLSEHNQSKOMKLLQKQSOVLQVQKHRLGHEHSAVLARSKLESCLRELQHNRSK 240
Db 197 ABLLSEHNQSKOMKLLQKQSOVLQVQKHRLGHEHSAVLARSKLESCLRELQHNRSK 256
QY 241 BEGVQARAEERKKEVTSHFQVTLNDIQLOMQHNRNKLQENMELAEKLLIEQY 300
Db 257 BEGVQARAEERKKEVTSHFQVTLNDIQLOMQHNRNKLQENMELAEKLLIEQY 316
QY 301 ELREEHIDKVPKHDLQOQLVDKLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
Db 317 ELREEHIDKVPKHDLQOQLVDKLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376
QY 361 KQOETHLQKQALYATEKFEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETTMYRSRW 420
Db 377 KQOETHLQKQALYATEKFEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETTMYRSRW 436
QY 421 ESSNKALLEWAEKTVRDKLEGLQVKIQRLKCRALQTERNDLNKRVQDLSAGGQSSL 480
Db 437 ESSNKALLEWAEKTVRDKLEGLQVKIQRLKCRALQTERNDLNKRVQDLSAGGQSSL 496
QY 481 TDSGPERRPGGAQAPSSPRVTEAPCYPGAPSTASGOTGPQETSARA 530
Db 497 TDSGPERRPGGAQAPSSPRVTEAPCYPGAPSTASGOTGPQETSARA 546

RESULT 4
Q86T85 PRELIMINARY; PRT; 546 AA.
ID Q86T85
AC Q86T85;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Hypothetical protein DKZp451J0118.
GN DKZP451J0118
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ansorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL; AL832637; CAD89952.1; -.
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61846 MW; 367183A3AA7B6C2 CRC64;

Query Match
Best Local Similarity 33.6%; Score 178; DB 4; Length 546;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 SQRMCELMKQOETHLQKQALYATEKFEFQNTLSKSEVFTTFQEMEKMTKKIKLEKE 412
Db 369 SQRMCELMKQOETHLQKQALYATEKFEFQNTLSKSEVFTTFQEMEKMTKKIKLEKE 428
```


Qy	413	TTMYRSRWSSSKALLEMAEKTVDKKEGLEGVKIQRLEKLCRALQTERNDLNKRVDLSA	474
Db	429	TTMYRSRWSSSKALLEMAEKTVDKKEGLEGVKIQRLEKLCRALQTERNDLNKRVDLSA	488
Qy	473	SAGGQSGLTDSGPERRPEGGGAQAPSPRVTEAPCYPCGAPSTASGQTGPQEPTSARA	530
Db	489	SAGGQSGLTDSGPERRPEGGGAQAPSPRVTEAPCYPCGAPSTASGQTGPQEPTSARA	546
 RESULT 5			
ID	Q86Y86	PRELIMINARY; PRT; 116 AA.	
AC	Q86Y86;		
DT	01-JUN-2003 (TrEMBLrel. 24, Created)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)		
DE	Hypothetical protein.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI_TaxID=9606;	[1]		
RN	SEQUENCE FROM N.A.		
RP	SEQUENCE		
RC	TISSUE=uterus;		
RA	Straussberg R.;		
RL	Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC046565; AAB46565.1; -		
KW	Hypothetical protein.		
SEQ	SEQUENCE 116 AA; 12623 MW; B99B79EAACAAB43F CRC64;		
 Query Match 21.9%; Score 116; DB 4; Length 116; Best Local Similarity 100.0%; Pred. No. 3e-111; Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0			
Qy	415	MYRSRWSSSKALLEMAEKTVDKKEGLEGVKIQRLEKLCRALQTERNDLNKRVDLSA	474
Db	1	MYRSRWSSSKALLEMAEKTVDKKEGLEGVKIQRLEKLCRALQTERNDLNKRVDLSA	60
Qy	475	GCGQSGLTDSGPERRPEGGGAQAPSPRVTEAPCYPCGAPSTASGQTGPQEPTSARA	530
Db	61	GCGQSGLTDSGPERRPEGGGAQAPSPRVTEAPCYPCGAPSTASGQTGPQEPTSARA	116
 RESULT 6			
ID	Q8C5K1	PRELIMINARY; PRT; 150 AA.	
AC	Q8C5K1		
DT	01-MAR-2003 (TrEMBLrel. 23, Created)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	Hypothetical protein (Fragment).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.		
NCBI_TaxID=10090;	[1]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=C57BL/6J; TISSUE=Olfactory brain;		
RC	MEDLINE=22354683; PubMed=1246851;		
RX	The FANTOM Consortium.		
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;		
RT	"Analysis of the mouse transcriptome based on functional annotation of		
RT	60,770 full-length cDNAs."		
RL	Nature 420:563-573 (2002).		
DR	EMBL; AK078192; BAC37168.1; -		
KW	Hypothetical protein.		
FT	NON TPR 150		
SEQ	SEQUENCE 150 AA; 16490 MW; C444717503BA42E49 CRC64;		
 Query Match 4.0%; Score 21; DB 11; Length 150; Best Local Similarity 100.0%; Pred. No. 6.8e-13; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0			

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 LNTLSTPEKLAALCKKYA 181
|||||
Db 97 LNTLSTPEKLAALCKKYA 115
|||||

RESULT 9

Q8BUK2 ID Q8BUK2 PRELIMINARY; PRT; 463 AA.

AC Q8BUK2; 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DE Muscle-derived protein MDP77 variant 1 (Fragment).

GN 231000IN14RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

R2 SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Heart;

RX MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs."

RL Nature 420:563-573(2002).

DR EMBL; AK084639; BAC39238.1; --

DR MGD; MGI:1916756; 231000IN14RIK.

FT NON TER 463 463

SQ SEQUENCE 463 AA; 54090 MW; 1B8EF81D73D58AA CRC64;

Query Match 3.6%; Score 19; DB 11; Length 463;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 AVLARKSLSCRELRQHN 236
|||||
Db 210 AVLARKSLSCRELRQHN 228
|||||

RESULT 10

Q8BHN1 ID Q8BHN1 PRELIMINARY; PRT; 524 AA.

AC Q8BHN1; 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Hypothetical protein.

GN RBP7.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

R2 SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Cortex, Head, and Testis;

RX MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs."

RL Nature 420:563-573(2002).

DR EMBL; AK030100; BAC26785.1; --

DR EMBL; AK031783; BAC27547.1; --

DR EMBL; AK044130; BAC31791.1; --

DR MGD; MGI:1194910; Rbp7.

DR GO; GO:0005634; C:nucleus; IDA.

DR GO; GO:0016564; F:transcriptional repressor activity; IDA.

DR GO; GO:0000122; P:negative regulation of transcription from P. . . ; IDA.

KW Hypothetical protein.

SQ SEQUENCE 524 AA; 60308 MW; 0228777633E4ED7C CRC64;

Query Match 3.6%; Score 19; DB 11; Length 524;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 LNTLSTPEKLAALCKKYA 181
|||||
Db 146 LNTLSTPEKLAALCKKYA 164
|||||

RESULT 11

Q9NUQ3 ID Q9NUQ3 PRELIMINARY; PRT; 528 AA.

AC Q9NUQ3; 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)

DE Hypothetical protein FLJ11209.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

R2 SEQUENCE FROM N.A.

RC TISSUE=Placenta;

RA Isogai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,

RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaika H., Kondo H., Sugawara M.,

RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,

RA Yatanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,

RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,

RA Ninomiya K., Iwayanagi T.;

RT "NEDO human cDNA sequencing project.";

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK002071; BAA92068.1; --

DR Genew; HGNC:18578; CXorf15.

KW Hypothetical protein.

SQ SEQUENCE 528 AA; 60605 MW; CA88D781DE06ACB3 CRC64;

Query Match 3.6%; Score 19; DB 4; Length 528;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 LNTLSTPEKLAALCKKYA 181
|||||
Db 146 LNTLSTPEKLAALCKKYA 164
|||||

RESULT 12

Q8VBT1 ID Q8VBT1 PRELIMINARY; PRT; 584 AA.

AC Q8VBT1; 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)

DE Muscle-derived protein MDP77 variant 2 (Muscle-derived protein MDP77

DE variant 1)

GN 231000IN14RIK OR MDP77.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

R2 SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Heart;

RX MEDLINE=21663722; PubMed=11805063;

RA Benson K.F., Chada K.;

RT "Molecular Characterization of the Mouse In(10)17Rk Inversion and

RT Identification of a Novel Muscle-Specific Gene at the Proximal

RT Breakpoint.";

RL Genetics 160:279-287(2002).

DR EMBL; AF422245; AAL33910.1; --

DR EMBL: AF422244; AAL33909.1; -.
DR MGI: 1916756; 2310001N14RIK.
SQ SEQUENCE 684 AA; 77049 MW; E8A64E28EF56E31 CRC64;
Query Match 3.6%; Score 19; DB 11; Length 684;
Best Local Similarity 100.0%; Pred. No. 3.1e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 218 AVLARSKLESCLRELQRHN 236
DB 209 AVLARSKLESCLRELQRHN 227
RESULT 13
Q9POX1 PRELIMINARY; PRT; 186 AA.
AC Q9POX1, 20, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Lipopolysaccharide specific response-5 protein (Fragment).
GN LSR5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chai Y.B., Zhao Z.L., Zhu P., Yan W., Chen N.C., Wang Q., Yue L.,
Chen S.M.,
RT "New Homo sapiens gene from dental pulp cells."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF143740; AAF70546.2; -.
FT NON TER 1
SQ SEQUENCE 186 AA; 21464 MW; 93BF8383096B0A CRC64;
Query Match 3.2%; Score 17; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 397 QEMKTKIKKLEKET 413
DB 38 QEMKTKIKKLEKET 54
RESULT 14
Q8N3S2 PRELIMINARY; PRT; 505 AA.
AC Q8N3S2, 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Hypothetical protein (Fragment).
GN DKFPZ451A175.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Koehrer X., Beyer A., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL832322; CAD38617.1; -.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 505 AA; 56957 MW; 1AA33548A2F1DADE CRC64;
Query Match 3.2%; Score 17; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 220 LARSKLESCLRELQRHN 236
DB 220 LARSKLESCLRELQRHN 236
RESULT 15
Q91969 PRELIMINARY; PRT; 676 AA.
AC Q91969, 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Muscle derived protein.
GN MDP77.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA MBLIME=20175243; PubMed=10708594;
RX Uyeda A., Fukui I., Fujimori K., Kiyosue K., Nishimune H., Kasai M.,
Taguchi T.;
RT "MDP77: A novel neurite-outgrowth-promoting protein predominantly
expressed in chick muscles."
RL Biochem. Biophys. Res. Commun. 269:564-569(2000).
DR EMBL: D89399; BAA94755.1; -.
DR PIR: JC7222; JC7222.
SQ SEQUENCE 676 AA; 77020 MW; FCEA9B393250EE94 CRC64;
Query Match 3.2%; Score 17; DB 13; Length 676;
Best Local Similarity 100.0%; Pred. No. 3.6e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 220 LARSKLESCLRELQRHN 236
DB 207 LARSKLESCLRELQRHN 223
RESULT 16
Q8N3L3 PRELIMINARY; PRT; 715 AA.
AC Q8N3L3, 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein DKFPZ451P022 (Fragment).
GN DKFPZ451P022.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
Csaranger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL834248; CAD38924.2; -.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 715 AA; 79798 MW; 11C50191BCD26582 CRC64;
Query Match 3.2%; Score 17; DB 4; Length 715;
Best Local Similarity 100.0%; Pred. No. 3.8e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 220 LARSKLESCLRELQRHN 236
DB 240 LARSKLESCLRELQRHN 256
RESULT 17
Q86T52

ID	Q86T52	PRELIMINARY;	PRT;	718 AA.	
AC	Q86T52;				
DT	01-JUN-2003 (TrEMBLrel. 24, Created)				
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)				
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)				
DE	Hypothetical protein DKFZp451G083 (Fragment).				
GN	DKFZP451G083.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Human skeletal muscle;				
RA	Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A.,				
RA	Fobo G., Han M., Wiemann S.;				
RL	Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.				
EMBL	AL831988; CAD91140.1; -.				
KM	Hypothetical protein.				
DR					
FT	NON TER				
SQ	SEQUENCE 718 AA; 80290 MW; 16C3BP22C19559A4 CRC64;				
	Query Match 3.2%; Score 17; DB 4; Length 718;				
	Best Local Similarity 100.0%; Pred. No. 3.9e-08;				
	Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	220 LARSKLSLCRELQHN 236				
DB	243 LARSKLSLCRELQHN 259				
RESULT 18					
Q8RXD7					
ID	Q8RXD7	PRELIMINARY;	PRT;	404 AA.	
AC	Q8RXD7;				
DT	01-JUN-2002 (TrEMBLrel. 21, Created)				
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	Hypothetical protein (At5G50840).				
GN	AT5G50840.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.				
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,				
RA	Palm C.J., Bowser L., Jones T., Banh J., Jones T., Kamiya A., Chen H.,				
RA	Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,				
RA	Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,				
RA	Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,				
RA	Ecker J., Theologis A., Davis R.W.;				
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.				
RP	[2]				
RP	SEQUENCE FROM N.A.				
RA	Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,				
RA	Dale J.M., Hayashizaki Y., Ishida J., Jones T., Kamiya A., Narusaka M.,				
RA	Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,				
RA	Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,				
RA	Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,				
RA	Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;				
RT	"Arabidopsis ORF clones";				
RL	Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AY081318; RAL91207.1; -.				
DR	EMBL; BT008857; AAP68296.1; -.				
KW	Hypothetical protein.				
SQ	SEQUENCE 404 AA; 46796 MW; 3C31C38A7E437DF6 CRC64;				
	Query Match 2.1%; Score 11; DB 10; Length 404;				
	Best Local Similarity 100.0%; Pred. No. 0.036;				
	Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	224 KLESICRELQ 234				
DB	163 KLESICRELQ 173				
RESULT 19					
Q8BZA4					
ID	Q8BZA4	PRELIMINARY;	PRT;	181 AA.	
AC	Q8BZA4;				
DT	01-JUN-2001 (TrEMBLrel. 17, Created)				
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)				
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)				
DE	Lipopolysaccharide-specific response 5-like protein (Hypothetical				
DE	protein DKFZp451G0616).				
GN	DKFZP451G0616.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Kuroda-Kawaguchi T., Skaletsky H., Minx P.J., Brown L.G., Rozen S.,				
RA	Wilson R.K., Waterston R.H., Page D.C.;				
RT	"The DNA sequence of the human Y chromosome";				
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Ansorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,				
RA	Osanger A., Wiemann S.;				
RL	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.				
EMBL	AF332225; AK13477.1; -.				
EMBL	AL832583; CAD89940.1; -.				
DR	Gene; HGNC:18577; Cyorf158.				
KW	Hypothetical protein.				
SQ	SEQUENCE 181 AA; 20858 MW; 4BA8B930933012A1 CRC64;				
	Query Match 1.9%; Score 10; DB 4; Length 181;				
	Best Local Similarity 100.0%; Pred. No. 0.19;				
	Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	403 TKKIKKLEKE 412				
DB	44 TKKIKKLEKE 53				
RESULT 20					
Q8BUB8					
ID	Q8BUB8	PRELIMINARY;	PRT;	190 AA.	
AC	Q8BUB8;				
DT	01-MAR-2003 (TrEMBLrel. 23, Created)				
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)				
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)				
DE	Muscle-derived protein MDP77 variant 1.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Heart;				
RX	MEDLINE=223354683; PubMed=12466851;				
RA	The FANTOM Consortium,				
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;				
RT	"Analysis of the mouse transcriptome based on functional annotation of				
RT	60,770 full-length cDNAs";				
RL	Nature 420:563-573(2002).				
DR	EMBL; AK084579; BAC39219.1; -.				
SQ	SEQUENCE 190 AA; 20852 MW; D96330563C9830D6 CRC64;				
	Query Match 1.9%; Score 10; DB 11; Length 190;				
	Best Local Similarity 100.0%; Pred. No. 0.2;				
	Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				

Qy 57 EELSRLQLEDI 66
 Db 60 EELSRLQLEDI 69

RESULT 21

Q8BDC1 PRELIMINARY; PRT; 637 AA.

AC Q8BDC1

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative ABC transporter, ATP-binding protein.

GN SWU.1431C.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1309;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=UAI59 / ATCC 700610 / Serotype C;

EX MEDLINE=22295063; PubMed=12397186;

RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,

RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,

RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;

RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental

RT pathogen.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

DR EMBL; AE014976; AAM59095.1; .

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005524; P:ATP binding; IEA.

DR GO; GO:0004009; P:ATP-binding cassette (ABC) transporter acti. . . ; IEA.

DR GO; GO:0000166; P:nucleotide binding; IEA.

DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR003593; AAA_ATPase.

DR Pfam; PF00005; ABC_tran; 2.

DR ProDom; PD000006; ABC_transporter; 2.

DR SMART; SW00382; AAA; 2.

DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.

DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.

XW ATP-binding; Complete proteome.

SQ SEQUENCE 637 AA; 72234 MW; CB75C5FB3B7716B37 CRC64;

Query Match 1.7%; Score 9; DB 16; Length 637;

Best Local Similarity 100.0%; Pred. No. 6.4;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 EELSRLQLED 65
 Db 629 EELSRLQLED 637

RESULT 22

Q9VPS3 PRELIMINARY; PRT; 826 AA.

AC Q9VPS3

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DE Putative ABC transporter, ATP-binding protein.

GN CG2839 protein.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkely;

EX MEDLINE=20196006; PubMed=107311132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

Sutton G.G., Wortman J.R., Yeandle M.D., Zhang Q., Chen L.X.,

Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champagne M., Pfeiffer B.D.,

Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

Balaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

Beeson K.Y., Benos P.V., Berman J.P., Bhandari D., Bolshakov S.,

Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

Glodek A., Gong P., Gottrell J.H., Gu Z., Guan P., Harris M.,

Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

Hosain D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,

Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

Zheng X.H., Zhong F., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,

Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*.";

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,

Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

Banson J., An H., Baldwin D., Banizon J., Beeson K.Y., Sussan D.A.,

Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,

Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,

Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Mostrefi A.,

McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuncio J.,

Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,

Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,

Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,

Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;

RT "Sequencing of *Drosophila melanogaster* genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,

Tupy J.L., Bergman C., Bernier B., Carlson J.W., Celniker S.E.,

Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,

Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,

Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield S.,

Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;

RT "Annotation of *Drosophila melanogaster* genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RA FlyBase;

RL Submitted (SEP-2002) to the EMBL/GenBank/DBSJ databases.

EMBL; AE003588; AAF51469.2; .

MEDLINE=22092622; PubMed=12097910;
Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.,
"Sequence and analysis of chromosome 2 of Dictyostelium discoideum.",
Nature 418:79-85(2002).
[2]
SEQUENCE FROM N.A.
STRAIN=AX4;
Baumgart C.;
Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
EMBL; AC116986; AA051852.1; -.
Hypothetical protein.
KW
SEQUENCE 113 AA; 12894 MW; 96243B5C6A3EDA1 CRC64;
[2]
Query Match 1.5%; Score 8; DB 5; Length 113;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 247 AREEEKR 254
|||||
DB 60 AREEEKR 67
|||||
RESULT 25
QBP61 PRELIMINARY; PRT; 135 AA.
ID QBP61;
AC QBP61;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Flagellar protein.
GN FLIO OR XAC1945.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
SEQUENCE FROM N.A.
STRAIN=306 / ATCC 13902 / XV 101;
MEDLINE=2022145; PubMed=1204217;
da Silva A.C.P., Ferro J.A., Reinach P.C., Farah C.S., Furlan L.R.,
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
Camarotte G., Canavari F., Cardoso J., Chamberg F., Chaplana L.P.,
Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Locall E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
Spindola L.A.F., Takita M.A., Tamura R.B., Teixeira B.C., Tezza R.I.D.,
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
Setubal J.C., Kitajima J.P.;
"Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities.",
Nature 417:459-463(2002).
RL Nature 417:459-463(2002).
DR EMBL; AE011830; BAM36807.1; -.
DR InterPro; IPR007442; FLIO.
DR Pfam; PF04347; FLIO; 1.
KW Complete proteome.
SQ
SEQUENCE 135 AA; 13836 MW; 38A5CFE543045981 CRC64;
Query Match 1.5%; Score 8; DB 16; Length 135;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 493 GAQAPSSP 500
|||||
DB 19 GAQAPSSP 26
|||||

DR EMBL; D88189; BAA13557.1; -.
 DR GO; GO:0019089; P:Viral transmission; IEA.
 DR InterPro; IPR007683; VapD_N.
 DR Pfam; PF04605; VapD_N; 1.
 DR SEQUENCE 143 AA; 16728 MW; 39C76D7F45834327 CRC64;
 SQ
 Query Match 1.5%; Score 8; DB 2; Length 143;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 EGVQVARE 249
 DB 87 EGVQVARE 94
 |||||
 |||||

RESULT 28
 017194 PRELIMINARY; PRT; 195 AA.
 ID 017194
 AC 017194
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-OCT-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein C08E3.4.
 GN C08E3.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA Waterston R.;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium."
 RL Science 282:2012-2018 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Miller N., Kramer J., Keppler D.;
 RT "The sequence of C. elegans cosmid C08E3.4";
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF025457; AAB70965.2; -.
 DR PIR; T32354; T32354.
 DR WormPep; C08E3.4; C832119.
 DR InterPro; IPR002900; DUF38.
 DR Pfam; PF01827; FTH; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 195 AA; 22534 MW; A750F3C8AD1BFA5F CRC64;
 Query Match 1.5%; Score 8; DB 5; Length 195;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 AERLKKLI 297
 DB 65 AERLKKLI 72
 |||||
 |||||

RESULT 29
 083C21 PRELIMINARY; PRT; 206 AA.
 ID 083C21
 AC 083C21
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein.
 GN CBU1314.

RESULT 26
 08P9E8 PRELIMINARY; PRT; 135 AA.
 ID 08P9E8
 AC 08P9E8
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Flagellar protein.
 GN FliO OR XCL1917.
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=340;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33913 / NCPPB 528;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach P.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardozo J., Chabergo F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.P., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado A.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.P.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C. de Souza R.F.,
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.P.,
 RA Setubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities."
 RL Nature 417:459-463 (2002).
 DR EMBL; AE012296; AAM41206.1; -.
 DR InterPro; IPR007442; FliO.
 DR Pfam; PF04347; FliO; 1.
 DR Complete proteome.
 SQ SEQUENCE 135 AA; 13849 MW; EF2AD52B981A518D CRC64;
 Query Match 1.5%; Score 8; DB 16; Length 135;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 493 GAQAPSSP 500
 DB 19 GAQAPSSP 26
 |||||
 |||||

RESULT 27
 P70717 PRELIMINARY; PRT; 143 AA.
 ID P70717
 AC P70717
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE VapD-homolog.
 OS Actinobacillus actinomycetemcomitans (Haemophilus
 OS actinomycetemcomitans).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Actinobacillus.
 OX NCBI_TaxID=714;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Y4;
 RX MEDLINE=97148607; PubMed=9020051;
 RA Yoshida Y., Nakano Y., Yamashita Y., Koga T.;
 RT "The gnd gene encoding a novel 6-phosphogluconate dehydrogenase and
 RT its adjacent region of Actinobacillus actinomycetemcomitans
 RT chromosomal DNA.";
 RL Biochem. Biophys. Res. Commun. 230:220-225 (1997).

CS Coccidia burnetii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
 OC Coccidiaceae; Coccidia.
 OX NCBI_TaxID=777;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nine Mile phase I / RSA 493;
 RC MEDLINE=22608657; PubMed=12704232;
 RA Seshadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
 RA Nelson W.C., Ward N.L., Tettelin H., Davidsen T.M., Beanan M.J.,
 RA DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
 RA Khouri H.M., Lee K.H., Carty H.A., Scanlan D., Heinzen R.A.,
 RA Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.F.;
 RT "Complete genome sequence of the Q-fever pathogen, Coccidia
 burnetii.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460 (2003).
 DR EMBL; AB016964; AAC090819.1; -;
 DR TIGR; CHU1314; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 206 AA; 23477 MW; 1C83C21054AED92D CRC64;
 Query Match 1.5%; Score 8; DB 16; Length 206;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 404 KKIKKLEK 411
 Db 69 KKIKKLEK 76
 RESULT 30
 Q9SCX3 PRELIMINARY; PRT; 224 AA.
 AC Q9SCX3;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Elongation factor 1B alpha-subunit.
 GN EF1BAPHAZ OR T20D1_30/AT5G19510.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20086877; PubMed=10618495;
 RA Hexicourt F., Jupin I.;
 RT "Molecular cloning and characterization of the Arabidopsis thaliana
 alpha-subunit of elongation factor 1B.";
 RL FEBS Lett. 464:148-152 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yamamura Y., Yu G.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene T20D1_30/AT5G19510.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Banno P., Dale J.M., Goldsmith A.D., Lee J.M.,
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
 RA Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,
 RA Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
 RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene T20D1_30/AT5G19510.";

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ249597; CAB64730.1; -;
 DR EMBL; AF360304; AAK26014.1; -;
 DR EMBL; AY056354; AAL07240.1; -;
 DR PIR; PA0110; PA0110.
 DR PIR; T52558; T52558.
 DR GO; GO:0005853; C:cytotoxic translation elongation factor 1.; IEA.
 DR GO; GO:0003746; F:translation elongation factor activity; IEA.
 DR GO; GO:0006414; P:translational elongation; IEA.
 DR InterPro; IPR001326; EFl_BD.
 DR InterPro; IPR004046; GST_Cterm.
 DR Pfam; PF00736; EFlBD; 1.
 DR Pfam; PF00043; GST_C; 1.
 DR PROSITE; PS00824; EFlBD_1; 1.
 DR PROSITE; PS00825; EFlBD_2; 1.
 KW Elongation factor.
 SQ SEQUENCE 224 AA; 24201 MW; F9B9F178A60CB3B4 CRC64;
 Query Match 1.5%; Score 8; DB 10; Length 224;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 21 AAPAVEAE 28
 Db 84 AAPAVEAE 91
 RESULT 31
 O06077 PRELIMINARY; PRT; 229 AA.
 ID O06077;
 AC O06077;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein (Probable membrane protein).
 GN ML2615 OR MLCU622.14.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RC MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Biglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus.";
 RL Nature 409:1007-1011 (2001).
 DR EMBL; Z95398; CAB08802.1; -;
 DR EMBL; AL583926; CAC32147.1; -;
 DR PIR; E87236; E87236.
 DR Leproma; ML2615; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 229 AA; 24416 MW; 9A839BF68B925E0A CRC64;
 Query Match 1.5%; Score 8; DB 16; Length 229;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 18 PSQAAPAV 25
 Db 200 PSQAAPAV 207
 RESULT 32
 Q8R9Z0

Q8R9Z0 PRELIMINARY; PRT; 255 AA.
Q8R9Z0;
AC 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 22, Last annotation update)
DE Flagellar biosynthesis/type III secretory pathway protein.
GN FLJH OR TT21440.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NE4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Huang Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EXBL; AB013102; AM24662.1; --
KW Complete Proteome.
SQ SEQUENCE 255 AA; 29724 MW; 910AS7784B8B859 CRC64;
Query Match 1.5%; Score 8; DB 16; Length 255;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 331 MLKAEER 338
Db |||||
75 MLKAEER 82
RESULT 33
Q8R9J9 PRELIMINARY; PRT; 255 AA.
AC Q8R9J9;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Type IV pili biogenesis protein PilP.
GN PILP OR PP0851
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.B., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., White O., Peterson J., Khouri H., Hance I.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzes A.,
RA Uterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuenmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440.";
RL Environ. Microbiol. 4:799-808(2002).
DR EMBL; AB016777; AAN66476.1; --
DR TIGR; PP0851; --
DR InterPro; IPR000437; Prock_lipoprot_S.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF00515; TPR; 4.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 255 AA; 28402 MW; C8FD220D181124FE CRC64;
Query Match 1.5%; Score 8; DB 16; Length 255;
Best Local Similarity 100.0%; Pred. No. 30;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 425 KALLEMAE 432
Db |||||
178 KALLEMAE 185
RESULT 34
Q86LQ1 PRELIMINARY; PRT; 264 AA.
AC Q86LQ1;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical muscle-derived protein.
OS Branchiostoma belcheri (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma
OX NCBI_TaxID=7741;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang H., Lin Y.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV191780; AAO45170.1; --
KW Hypothetical protein.
SQ SEQUENCE 264 AA; 31199 MW; EAOB82C96708815B CRC64;
Query Match 1.5%; Score 8; DB 5; Length 264;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 373 LYTERFEE 380
Db |||||
92 LYTERFEE 99
RESULT 35
Q44174 PRELIMINARY; PRT; 281 AA.
AC Q44174;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE F58P6.1 protein.
GN F58P6.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Dante M., Kramer J., Gibson A.;
RT "The sequence of C. elegans cosmid F58P6.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF036699; AAB88358.1; --
DR PIR; C88638; C88638.
DR WormPep; F58P6.1; CB17136.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.

DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR InterPro; IPR001150; Form_actrans_GR.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF01484; Col_cuticle_N; 1.
DR PROSITE; PS00850; GLY_RADICAL; 1.
SQ SEQUENCE 281 AA; 27995 MW; EA2112F6CF753419 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 281;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 GGPGEDEGA 84
Db 241 GGPGEDEGA 248
|||||
|||

RESULT 36
Q9SZW2 PRELIMINARY; PRT; 312 AA.
ID Q9SZW2
AC Q9SZW2
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN F6G3.120 OR A74G30090.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
EK [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078464; CAB43844.1; -;
DR EMBL; AL161576; CAB81002.1; -;
DR PIR; T08985; T08985
KW Hypothetical protein.
SQ SEQUENCE 312 AA; 35914 MW; D169444734E31331 CRC64;

Query Match 1.5%; Score 8; DB 10; Length 312;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 QQAQEMLK 333
Db 240 QQAQEMLK 247
|||||
|||

RESULT 37
Q8SA04 PRELIMINARY; PRT; 324 AA.
ID Q8SA04
AC Q8SA04;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to root cap protein.
GN QJ1656 All.6.
OS Oryza sativa (japonica cultivar-group).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OC NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
clone:QJ1656 All.6";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003448; BAB85317.1; -;
DR Gramene; Q8SA04; -;
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000515; BPD transp.
DR PROSITE; PS00402; BPD TRANSP INN MEMBER; 1.
SQ SEQUENCE 324 AA; 34945 MW; 0F37890FD24EFF7 CRC64;

Query Match 1.5%; Score 8; DB 10; Length 324;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AQAQSSPR 501
Db 22 AQAQSSPR 29
|||||
|||

RESULT 38
Q8ELJ3 PRELIMINARY; PRT; 346 AA.
ID Q8ELJ3
AC Q8ELJ3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical conserved protein.
GN OB3234.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OC NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HT831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AF004604; BAC15190.1; -;
DR GO; GO:0030288; C:periplasmic space (sensu Gram-negative Bact...; IEA.
DR InterPro; IPR000437; Prok_lipoprot_s.
DR InterPro; IPR005064; UPF0065.
DR Pfam; PF03401; UPF0065; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 346 AA; 37487 MW; 12B9A69D4807025A CRC64;

Query Match 1.5%; Score 8; DB 16; Length 346;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 379 EEPQNTLS 386
Db 291 EEPQNTLS 298
|||||
|||

RESULT 39
Q9Q1U7 PRELIMINARY; PRT; 348 AA.
ID Q9Q1U7
AC Q9Q1U7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR InterPro; IPR001150; Form_actrans_GR.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF01484; Col_cuticle_N; 1.
DR PROSITE; PS00850; GLY_RADICAL; 1.
SQ SEQUENCE 281 AA; 27995 MW; EA2112F6CF753419 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 281;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 GGPGEDEGA 84
Db 241 GGPGEDEGA 248
|||||
|||

RESULT 36
Q9SZW2 PRELIMINARY; PRT; 312 AA.
ID Q9SZW2
AC Q9SZW2
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN F6G3.120 OR A74G30090.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
EK [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078464; CAB43844.1; -;
DR EMBL; AL161576; CAB81002.1; -;
DR PIR; T08985; T08985
KW Hypothetical protein.
SQ SEQUENCE 312 AA; 35914 MW; D169444734E31331 CRC64;

Query Match 1.5%; Score 8; DB 10; Length 312;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 QQAQEMLK 333
Db 240 QQAQEMLK 247
|||||
|||

RESULT 37
Q8SA04 PRELIMINARY; PRT; 324 AA.
ID Q8SA04
AC Q8SA04;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to root cap protein.
GN QJ1656 All.6.
OS Oryza sativa (japonica cultivar-group).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OC NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
clone:QJ1656 All.6";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003448; BAB85317.1; -;
DR Gramene; Q8SA04; -;
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000515; BPD transp.
DR PROSITE; PS00402; BPD TRANSP INN MEMBER; 1.
SQ SEQUENCE 324 AA; 34945 MW; 0F37890FD24EFF7 CRC64;

Query Match 1.5%; Score 8; DB 10; Length 324;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AQAQSSPR 501
Db 22 AQAQSSPR 29
|||||
|||

RESULT 38
Q8ELJ3 PRELIMINARY; PRT; 346 AA.
ID Q8ELJ3
AC Q8ELJ3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical conserved protein.
GN OB3234.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OC NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HT831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AF004604; BAC15190.1; -;
DR GO; GO:0030288; C:periplasmic space (sensu Gram-negative Bact...; IEA.
DR InterPro; IPR000437; Prok_lipoprot_s.
DR InterPro; IPR005064; UPF0065.
DR Pfam; PF03401; UPF0065; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 346 AA; 37487 MW; 12B9A69D4807025A CRC64;

Query Match 1.5%; Score 8; DB 16; Length 346;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 379 EEPQNTLS 386
Db 291 EEPQNTLS 298
|||||
|||

RESULT 39
Q9Q1U7 PRELIMINARY; PRT; 348 AA.
ID Q9Q1U7
AC Q9Q1U7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ORF50b protein short form (Hypothetical protein).
 OS Saimiriine herpesvirus 2.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=10381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C488;
 RX MEDLINE=20149233; PubMed=10683339;
 RA Thurau M., Whitehouse A., Wittmann S., Fickenscher H.;
 RT "Distinct transcriptional and functional properties of the R
 transactivator gene, orf50, of the transforming herpesvirus saimiri
 strain C488.";
 RL Virology 288:167-177 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C5753;
 RA Ensser A., Thurau M., Wittmann S., Fickenscher H.;
 RT "Genomic sequence of herpesvirus saimiri C488 which is capable of
 transforming human T cells.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C488;
 RA Fickenscher H., Thurau M., Wittmann S., Ensser A.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ131935; CAB65421.1; -;
 DR EMBL; AJ410487; CAC84289.1; -;
 DR EMBL; AJ410493; CAC84346.1; -;
 DR GO; GO:0004386; F:Helicase activity; IEA.
 DR GO; GO:0009233; F:Peptidase activity; IEA.
 DR GO; GO:0016563; F:Transcriptional activator activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR004998; Herpes_TAF50.
 DR PFam; PF03326; Herpes_TAF50; 1.
 KW Hypothetical protein; Alternation: splicing; Helicase; Protease.
 SQ SEQUENCE 348 AA; 38472 MW; 8678581FAC735E8 CRC64;

Query Match 1.5%; Score 8; DB 12; Length 348;
 Best Local Similarity 100.0%; Pred.No. 40;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GLGKEITL 158
 DB 2 GLGKEITL 9

RESULT 40
 Q805G5 PRELIMINARY; PRT; 348 AA.
 AC Q805G5;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Saimiriine herpesvirus 2.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=10381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C5746, C5945, C5947, C5952, and C6661;
 RA Ensser A., Thurau M., Wittmann S., Fickenscher H.;
 RT "Genomic sequence of herpesvirus saimiri C488 which is capable of
 transforming human T cells.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ410486; CAC85015.1; -;
 DR EMBL; AJ410488; CAC85021.1; -;
 DR EMBL; AJ410489; CAC85027.1; -;
 DR EMBL; AJ410490; CAC85033.1; -;
 DR EMBL; AJ410491; CAC85039.1; -;

DR GO; GO:0016563; P:transcriptional activator activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR004998; Herpes_TAF50.
 DR PFam; PF03326; Herpes_TAF50; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 348 AA; 38449 MW; F512C3813CEID026 CRC64;
 Query Match 1.5%; Score 8; DB 12; Length 348;
 Best Local Similarity 100.0%; Pred.No. 40;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GLGKEITL 158
 DB 2 GLGKEITL 9

RESULT 42
 Q80BKO PRELIMINARY; PRT; 405 AA.
 AC Q80BKO;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Saimiriine herpesvirus 2.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=10381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A5747;
 RA Ensser A., Thurau M., Wittmann S., Fickenscher H.;
 RT "Genomic sequence of herpesvirus saimiri C488 which is capable of
 transforming human T cells.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ410482; CAC84997.1; -;

```

DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0016563; F:transcriptional activator activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR000637; AT_hoek.
DR InterPro: IPR004998; HePpes_TAF50.
DR Pfam: PF03326; Herpes_TAF50; 1.
DR PROSITE: PS00354; HMG_Y; 1.
KW Hypothetical protein.
SQ SEQUENCE 405 AA; 44720 MW; 29CC90C991B61572 CRC64;

Query Match 1.5%; Score 8; DB 12; Length 405;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GLGKEITL 158
Db 57 GLGKEITL 64

RESULT 43
Q805M7 PRELIMINARY; PRT; 405 AA.
ID Q805M7
AC Q805M7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Saimirine herpesvirus 2.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A6355, and A6051;
RA Essner A., Thraut M., Wittmann S., Pickenscher H.;
RT "Genomic sequence of herpesvirus saimiri C488 which is capable of
transforming human T cells.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ410484; CAC84283.1; -.
DR EMBL: AJ410483; CAC85003.1; -.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0016563; F:transcriptional activator activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR000637; AT_hoek.
DR InterPro: IPR004998; HePpes_TAF50.
DR Pfam: PF03326; Herpes_TAF50; 1.
DR PROSITE: PS00354; HMG_Y; 1.
KW Hypothetical protein.
SQ SEQUENCE 405 AA; 44783 MW; 7512ED30F51045B7 CRC64;

Query Match 1.5%; Score 8; DB 12; Length 405;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GLGKEITL 158
Db 57 GLGKEITL 64

RESULT 44
Q82DR9 PRELIMINARY; PRT; 413 AA.
ID Q82DR9
AC Q82DR9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SA44899.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.

```

```

OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL: AP005040; BAC72611.1; -.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR InterPro: IPR006680; Amidohydro_1.
DR Pfam: PF01979; Amidohydro_1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 413 AA; 41637 MW; 3B5CB2B80F14BDC0 CRC64;

Query Match 1.5%; Score 8; DB 16; Length 413;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 SGALRDVS 56
Db 297 SGALRDVS 304

RESULT 45
Q9L0M6 PRELIMINARY; PRT; 434 AA.
ID Q9L0M6
AC Q9L0M6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative hydrolase.
GN SCO4634 OR SCO82.05C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL939120; CAB77408.1; -.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR InterPro: IPR006680; Amidohydro_1.
DR Pfam: PF01979; Amidohydro_1; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 434 AA; 43748 MW; P2BA6B51A465AB38 CRC64;

```

Query Match 1.5%; Score 8; DB 16; Length 434;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 SGALRDVS 56
Db 303 SGALRDVS 310
|||||

RESULT 46
Q84VE3 PRELIMINARY; PRT; 436 AA.
AC Q84VE3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Muscle derived-like protein.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39347;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Cooper B., Hutchinson D., Park S., Guimil S., Luginbuhl P., Ellero C.,
RA Goff S., Glazebrook J.;
RT "Identification of Rice (Oryza sativa) Proteins Linked to the Cyclin-
RT Mediated Regulation of the Cell Cycle.";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY224541; AA072661.1; -. 9180958P88E2328 CRC64;
SQ SEQUENCE 436 AA; 48148 MW; 9180958P88E2328 CRC64;

Query Match 1.5%; Score 8; DB 10; Length 436;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 KLESICRE 231
Db 186 KLESICRE 193
|||||

RESULT 47
Q80BH6 PRELIMINARY; PRT; 439 AA.
AC Q80BH6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Saimirine herpesvirus 2.
OC Viruses; GSDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS907;
RA Ensser A., Thurau M., Wittmann S., Fickenscher H.;
RT "Genomic sequence of herpesvirus saimiri C488 which is capable of
RT transforming human T cells.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ410492; CAC85045.1; -.
DR GO; GO:0016563; F:transcriptional activator activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR004998; Herpes_TAF50.
DR Pfam; PF03326; Herpes_TAF50; 1.
KW Hypothetical protein.
SQ SEQUENCE 439 AA; 49007 MW; 2643B9384E505B2D CRC64;

Query Match 1.5%; Score 8; DB 12; Length 439;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 GLGKEITL 158
Db 93 GLGKEITL 100
|||||

RESULT 48
Q9ZLGS PRELIMINARY; PRT; 450 AA.
AC Q9ZLGS;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative.
GN JHP0612.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., Kirg B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT Gastric pathogen Helicobacter Pylori.";
RL Nature 387:176-180(1999).
DR EMBL; AE001493; AAD06193.1; -.
DR PIR; E71909; E71909.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR007560; Mr_Cat.
DR Pfam; PF04471; Mr_Cat; 1.
DR SMART; SM00487; DEXDc; 1.
KW Complete proteome.
SQ SEQUENCE 450 AA; 51347 MW; E15CFAFD2B9074C8 CRC64;

Query Match 1.5%; Score 8; DB 16; Length 450;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 63 LEDILSTY 70
Db 283 LEDILSTY 290
|||||

RESULT 49
Q8JZS6 PRELIMINARY; PRT; 451 AA.
AC Q8JZS6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Gibberella fujikuroi (Bakanae and foot rot disease fungus) (Fusarium moniliforme).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=5127;
RN [1]
RP SEQUENCE FROM N.A.
RA Yoshida H., Iizuka M., Hirabayashi T.;
RT "Cloning and sequencing of a cDNA encoding caldesmon-like protein of
RT Gibberella fujikuroi.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB02343; BAC16636.1; -.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 451 AA; 53407 MW; 0D2796E91E55D610 CRC64;

```
Query Match 1.5%; Score 8; DB 3; Length 451;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 REEEERK 255
Db 15 REEEERK 22

RESULT 50
Q8NMB0 PRELIMINARY; PRT; 484 AA.
AC Q8NMB0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NAD-dependent aldehyde dehydrogenases (EC 1.2.1.-).
CGL2668.
GN Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RL "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032."
DR EMBL: AP005282; BAC00062.1; -.
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR02086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
DR Oxidoreductase; Complete proteome.
KM OXIDOREDUCTASE; Complete proteome.
SQ SEQUENCE 484 AA; 51084 MW; A2E416EB585775F CRC64;

Query Match 1.5%; Score 8; DB 16; Length 484;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 AELEEHR 188
Db 80 AELEEHR 87

RESULT 51
Q8PEY9 PRELIMINARY; PRT; 501 AA.
AC Q8PEY9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Polyvinylalcohol dehydrogenase.
GN XAC4199.
OC Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Quaggio R.B., Monteiro-Vicorello C.B., van Sluys M.A., Almeida N.P.,
RA Canarotte G., Cannavan F., Cardoso J., Chamargo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locall E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.N., Miyaki C.Y., Moon D.H.,

Query Match 1.5%; Score 8; DB 12; Length 520;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GLGKEITL 158
Db 174 GLGKEITL 181

RESULT 53
Q8VCQ8 PRELIMINARY; PRT; 530 AA.
ID Q8VCQ8
```

```
RA Moreira L.M., Novo M.T.M., Okura V.X., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spínola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezsa R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.W., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL: AE012071; BAM39034.1; -.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002372; Bac_PQQ_repeat.
DR Pfam; PF01011; PQQ; 2.
KW Complete proteome.
SQ SEQUENCE 501 AA; 52817 MW; F21C9784E3652129 CRC64;

Query Match 1.5%; Score 8; DB 16; Length 501;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 AGQGSLT 481
Db 417 AGQGSLT 424

RESULT 52
Q9QIU8 PRELIMINARY; PRT; 520 AA.
ID Q9QIU8;
AC Q9QIU8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ORF50a protein long form (transcriptional regulator).
OS Salmirine herpesvirus 2.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C488;
RX MEDLINE=20149233; PubMed=10683339;
RA Thuraux M., Whitehouse A., Wittmann S., Meredith D., Fickenscher H.;
RT "Distinct transcriptional and functional properties of the R
RT transactivator gene, orf50, of the transforming herpesvirus salmiri
RT strain C488."
RL Virology 268:167-177(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C488;
RA Fickenscher H., Thuraux M., Wittmann S., Ensser A.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ131935; CAB65420.1; -.
DR EMBL: AJ410493; CAC84344.1; -.
DR GO; GO:0004386; P:phenylalanine activity; IEA.
DR GO; GO:0008233; P:peptidase activity; IEA.
DR GO; GO:0016563; P:regulation of transcription; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR004998; Herpes_TAF50.
DR Pfam; PF03326; Herpes_TAF50.1.
KW Alternative splicing; Helicase; Protease.
SQ SEQUENCE 520 AA; 58309 MW; 33CADF47EF50AD5D CRC64;

Query Match 1.5%; Score 8; DB 12; Length 520;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GLGKEITL 158
Db 174 GLGKEITL 181
```


AC Q8VC08;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to caldesmon 1.
GN CALDI OR 4833423D12RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019435; AAI19435.1; -.
DR MGD; MGI:88250; Caldi.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0005516; F:calmodulin binding; IEA.
DR GO; GO:0017022; F:myosin binding; IEA.
DR GO; GO:0005936; F:muscle contraction; IEA.
DR InterPro; IPR006017; Caldesmon.
DR InterPro; IPR006018; Caldesmon_LSP.
DR Pfam; PF02029; Caldesmon; 1.
DR PRINTS; PR01076; CALDESMON.
SQ SEQUENCE 530 AA; 50453 MW; 697BFACF8FD68A2E CRC64;
Query Match 1.5%; Score 8; DB 11; Length 530;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 247 AREEEKR 254
DB 333 AREEEKR 340
RESULT 54
Q8P3J1
ID Q8P3J1 PRELIMINARY; PRT; 539 AA.
AC Q8P3J1
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Polyvinylalcohol dehydrogenase.
GN XCC4080.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.P.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Canavan F., Cardoso J., Chamberg F., Chapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.P.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.P.,
RA "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463 (2002).
DR EMBL; AR012531; AAM43301.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0004812; F:RNA ligase activity; IEA.
DR GO; GO:0006418; P:amino acid activation; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002372; Bac_PQQ_repeat.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF01011; PQQ; 2
DR PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.
KW Complete proteome.
SQ SEQUENCE 539 AA; 56714 MW; A47CA59B5F43BF12 CRC64;
Query Match 1.5%; Score 8; DB 16; Length 539;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 474 AGGQSLT 481
DB 455 AGGQSLT 462
RESULT 55
Q93ZM4
ID Q93ZM4 PRELIMINARY; PRT; 558 AA.
AC Q93ZM4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE AT3G48870/T21J18.140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY056787; AAL10478.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_centr.
DR InterPro; IPR001270; Chaperin_c1pA/B.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF02861; Clp_N; 3.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00870; ClpAB_1; 1.
KW ATP-binding.
SQ SEQUENCE 558 AA; 61686 MW; 7B96AB5B1B57221 CRC64;
Query Match 1.5%; Score 8; DB 10; Length 558;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 127 BEIRQSD 134
DB 381 BEIRQSD 388
RESULT 56
Q97WX9
ID Q97WX9 PRELIMINARY; PRT; 568 AA.
AC Q97WX9;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)

01-OCT-2001 (TREMELrel. 18, Last sequence update)
 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 Hypothetical protein, plasmid pNOB8 orfs 620 and 630A
 homolog.
 SSO1981.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332726; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 Aways M.J., Chan-Weher C.-Y., Clausen I.G., Curtis B.A.,
 De Moors A., Eraso G., Fletcher C., Gordon P.M.K.,
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 Thi-Ngoc H.P., Redder P., Schenk M.B., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.P., Duguet M., Gaasterland T.,
 Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.,
 RA "the complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 DR EMBL; AE006805; AAK42172.1; -;
 DR FIR; E90364; E90364.
 DR InterPro; IPR003016; Lipovyl BS.
 DR PROSITE; PS00189; LIPOVL; 1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 568 AA; 63993 MW; D32E40D02D097985 CRC64;
 SQ
 Query Match 1.5%; Score 8; DB 17; Length 568;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 404 KKIKLEK 411
 DB 91 KKIKLEK 98
 RESULT 57
 Q8DJCO PRELIMINARY; PRT; 609 AA.
 ID Q8DJCO
 AC Q8DJCO
 DT 01-MAR-2003 (TREMELrel. 23, Created)
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
 DE Precorrin methylase.
 GN COBJ OR TLL1307.
 OS Synechococcus elongatus (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 OX NCBI_TaxID=32046;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=BP-1;
 RX MEDLINE=4225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsumoto M., Nakazaki N.,
 RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.,
 RA "Complete genome structure of the thermophilic cyanobacterium
 Thermosynechococcus elongatus BP-1.";
 RL DNA Res. 9:123-130(2002).
 DR EMBL; AP005373; BAC08859.1; -;
 DR GO; GO:0008168; F:methyltransferase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR GO; GO:0009236; P:vitamin B12 biosynthesis; IEA.
 DR InterPro; IPR002750; CbiG.
 DR InterPro; IPR006363; CobiG.
 DR InterPro; IPR000878; Cor/por_Mettransf.
 DR Pfam; PF01890; CbiG; 1.
 DR Pfam; PF00590; TP_methylase; 1.
 DR TIGRFAMs; TIGR01466; cobJ_cbiH; 1.
 KW Methyltransferase; Complete proteome.
 SQ SEQUENCE 609 AA; 54732 MW; 2A26F93FB24B920D CRC64;

Query Match 1.5%; Score 8; DB 16; Length 609;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 317 QQQLVDK 324
 DB 532 QQQLVDK 539
 RESULT 58
 Q98P27 PRELIMINARY; PRT; 649 AA.
 ID Q98P27
 AC Q98P27
 DT 01-OCT-2001 (TREMELrel. 18, Created)
 DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
 DE Hypothetical protein mli19638.
 GN MLI9638.
 OS Rhizobium loti (Mesorhizobium loti).
 OG Plasmid pMLB.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Ideawara K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003017; BAB54828.1; -;
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 KW Plasmid; Hypothetical protein; Complete proteome.
 SQ SEQUENCE 649 AA; 74481 MW; D8D49A4A5B52998 CRC64;
 Query Match 1.5%; Score 8; DB 16; Length 649;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 289 LAERLKKL 296
 DB 124 LAERLKKL 131
 RESULT 59
 Q92QD4 PRELIMINARY; PRT; 649 AA.
 ID Q92QD4
 AC Q92QD4
 DT 01-DEC-2001 (TREMELrel. 19, Created)
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)
 DE Hypothetical protein R01397.
 GN R01397 OR SMC01267.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=1021;
 RX MEDLINE=21596507; PubMed=11481430;
 RA Capela D., Barloy-Hubier F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.,
 RA "Analysis of the chromosome sequence of the legume symbiont

ATP-dependent Clp protease, ATP-binding subunit
(AT5950920/K3X7.7).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Columbia;
RX MEDLINE=99397451; PubMed=10470850;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA Miyajima N., Tabata S.
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
RT Sequence features of the regions of 1,011,550 bp covered by seventeen
RT P1 and TAC clones."
RL DNA Res. 6:183-195(1999).
RN [2]
RN SEQUENCE FROM N.A.
RP Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Zeng J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017063; SAB08738.1; -;
DR EMBL; AY102125; AAM26692.1; -;
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004518; F:nuclease activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0006289; F:nucleotide-excision repair; IEA.
DR InterPro; IPR003959; AAA ATPase centr.
DR InterPro; IPR001270; Chaprinin_clpA/B.
DR InterPro; IPR004176; Clp N.
DR InterPro; IPR001943; UvrE/C.
DR Pfam; PF00004; AAA; 2.
DR Pfam; PF02861; Clp N; 2.
DR PRINTS; PR00300; CLPPTASEA.
DR PROSITE; PS00870; CLPAB 1; 1.
DR PROSITE; PS00871; CLPAB 2; 1.
DR PROSITE; PS0151; UVR; 1.
KW ATP-binding; Protease.
SQ SEQUENCE 929 AA; 103452 MW; 438DEA514125F0BF CRC64;
Query Match 1.5%; Score 8; DB 10; Length 929;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 127 EIRQSD 134
DB 360 EIRQSD 367
RESULT 62
QUP82
ID QUP82 PRELIMINARY; PRT; 944 AA.
AC QUP82;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Mucin 4 (Fragment).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591787; CAC45976.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 649 AA; 74299 MW; 8578F0EA569721D5 CRC64;
Query Match 1.5%; Score 8; DB 16; Length 649;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 289 LAERLKKL 296
DB 124 LAERLKKL 131
RESULT 60
O48931
ID O48931 PRELIMINARY; PRT; 928 AA.
AC O48931;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ClpC.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RP Clarke A.K.;
RT "A cDNA clone coding for the chloroplast ClpC protein from Arabidopsis
RT thaliana";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF022909; AAC04687.1; -;
DR PIR; T52292; T52292.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004518; F:nuclease activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006289; F:nucleotide-excision repair; IEA.
DR InterPro; IPR003959; AAA ATPase.
DR InterPro; IPR001270; Chaprinin_clpA/B.
DR InterPro; IPR004176; Clp N.
DR InterPro; IPR001943; UvrE/C.
DR Pfam; PF00004; AAA; 2.
DR Pfam; PF02861; Clp N; 2.
DR PRINTS; PR00300; CLPPTASEA.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00870; CLPAB 1; 1.
DR PROSITE; PS0151; UVR; 1.
KW ATP-binding.
SQ SEQUENCE 928 AA; 103455 MW; F270B210C0D21C1D CRC64;
Query Match 1.5%; Score 8; DB 10; Length 928;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 127 EIRQSD 134
DB 359 EIRQSD 366
RESULT 61
Q9FI56
ID Q9FI56 PRELIMINARY; PRT; 929 AA.
AC Q9FI56;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

[1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=99077625; PubMed=9858486;
 RX Gibson I.K., Spurr-Michaud S., Moccia R., Zhan Q., Toribara N.,
 RA Ho S.B., Gargiulo A.R., Hill J.A. III;
 RA "MUGA and MUC5B transcripts are the prevalent
 RT ribonucleic acids of the human endocervix.";
 RRL Biol. Reprod. 60:58-64(1999).
 DR EMBL; AF058803; AAC34750.1; -.
 DR EMBL; AF058803; AAC34750.1; -.
 FT NON_TER 1 1
 FT NON_TER 944 944
 FT NON_TER 944 944
 SQ SEQUENCE 944 AA; 587CE9EC5085595 CRC64;

 Query Match 1.5%; Score 8; DB 4; Length 944;
 Best Local Similarity 100.0%; Pred.No. 99;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 513 STEASQOT 520
 DB 697 STEASQOT 704

 RESULT 63
 Q9M2Z6 PRELIMINARY; PRT; 952 AA.
 ID Q9M2Z6
 AC Q9M2Z6;
 DT 01-OCT-2000 (TREMELrel. 15, Created)
 DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE ATClpC.
 GS T21J18.140.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
 CX NCBI_TaxID=3702;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,
 RA Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X., Quetier F.,
 RA Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RP EU Arabidopsis sequencing project;
 RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AL12963; CAB87915.1; -.
 DR PIR; T49283; T49283.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003754; F:chaperone activity; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0004518; F:nuclease activity; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006289; F:nucleotide-excision repair; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003959; AAA_ATPase_centr.
 DR InterPro; IPR001270; Chaptin_c1pA/B.
 DR InterPro; IPR004176; Clp_N.
 DR InterPro; IPR001943; UvrE/C.
 DR Pfam; PF00004; AAA; 2.
 DR Pfam; PF02861; Clp_N; 3.
 DR Pfam; PF02151; UVR; 1.
 DR SMART; PRO0300; CLPPTHEASEA.
 DR PRINTS; SM00382; AAA; 2.
 DR PROSITE; PS00870; CLPAB_1; 1.
 DR PROSITE; PS00871; CLPAB_2; 1.
 DR PROSITE; PS00151; UVR; 1.
 KX ATP-binding
 SQ SEQUENCE 952 AA; 105771 MW; 295A0B61319E758C CRC64;

 Query Match 1.5%; Score 8; DB 10; Length 952;
 Best Local Similarity 100.0%; Pred.No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OX NCBI_TaxID=32025;
RN [1]_TaxID=32025;
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51449 / 3B1;
RX MEDLINE=22709201; PubMed=12810954;
RA Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
RA Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
RT "The complete genome sequence of the carcinogenic bacterium
RT Helicobacter hepaticus.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
DR EMBL; AF017145; AAP77222.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1021 AA; 112011 MW; 41ADF6BFI744F13C CRC64;
Query Match 1.5%; Score 8; DB 16; Length 1021;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 388 SSEVFTTF 395
DB 698 SSEVFTTF 705
|||||
RESULT 66
O76037 PRELIMINARY; PRT; 1045 AA.
AC O76037;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Mucin precursor (Fragment).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon mucosa;
RX MEDLINE=98285542; PubMed=9620877;
RA Nollet S., Moniaux N., Maury J.P., Petitprez D., Degand P., Laine A.,
RA Porchet N., Aubert J.P.;
RT "Human Mucin Gene MUC4: Organization of its 5' Region and of its
RT Polymorphic Central Tandem Repeat Array.";
RL Biochem. J. 332:739-748(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon mucosa;
RX MEDLINE=20381033; PubMed=10920259;
RA Choudhury A., Moniaux N., Wimpenny J.P., Hollingsworth M.A.,
RA Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin cDNA and its variants in pancreatic carcinoma.";
RL J. Biochem. 128:233-243(2000).
DR EMBL; AJ000281; CAA03985.1; --
DR GO; GO:0005201; Extracellular matrix structural constituent; NAS.
KW Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 >1045 MUCIN.
FT NON TER 1045 1045
SQ SEQUENCE 1045 AA; 106737 MW; 26F12DF2BE6D60A0 CRC64;
Query Match 1.5%; Score 8; DB 4; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 513 STEASGQT 520
DB 783 STEASGQT 790
|||||
RESULT 67

Q7UYN4 PRELIMINARY; PRT; 1101 AA.
AC Q7UYN4;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Similar to FKBP-rapamycin associated protein.
GN RB479.
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=1117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR EMBL; BX294133; CAD71608.1; --
KW Complete proteome.
SQ SEQUENCE 1101 AA; 123359 MW; AB8099B30867DD6A CRC64;
Query Match 1.5%; Score 8; DB 16; Length 1101;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 468 RVQDLASG 475
DB 450 RVQDLASG 457
|||||
RESULT 68
O9H4D6 PRELIMINARY; PRT; 1107 AA.
AC O9H4D6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE MUC4 protein splice variant svll.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Choudhury A., Moniaux N., Ringel J., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; A0277505; CAC14585.1; --
SQ SEQUENCE 1107 AA; 113576 MW; 2669ADF5DA53721 CRC64;
Query Match 1.5%; Score 8; DB 4; Length 1107;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 513 STEASGQT 520
DB 783 STEASGQT 790
|||||
RESULT 69
O9GZM2 PRELIMINARY; PRT; 1157 AA.
AC O9GZM2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE MUC4 protein splice variant svl3 (MUC4 protein splice variant

```

DE sv12).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ400850; CAC14135.1; -.
DR EMBL; AJ400849; CAC14134.1; -.
SQ SEQUENCE 1157 AA; 118925 MW; F46F039320969A9C CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1157;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASQGT 520
Db 783 STEASQGT 790

RESULT 70
Q9GZV6 PRELIMINARY; PRT; 1187 AA.
AC Q9GZV6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MUC4 protein splice variant sv14 (MUC4 protein variant VII).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20381033; PubMed=10920259;
RA Choudhury A., Moniaux N., Wimpenny J.P., Hollingsworth M.A.,
RA Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin cDNA and its variants in pancreatic carcinoma.";
RL J. Biochem. 128:233-243(2000).
DR EMBL; AJ400851; CAC14136.1; -.
DR EMBL; AJ400633; CAC10062.1; -.
DR GO; GO:0016020; C:membrane; NAS.
DR GO; GO:0005201; P:extracellular matrix structural constituent; NAS.
SQ SEQUENCE 1187 AA; 121925 MW; 62E4C93F86443B8F CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1187;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASQGT 520
Db 783 STEASQGT 790

RESULT 71
Q9H487 PRELIMINARY; PRT; 1199 AA.
AC Q9H487;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

```

```

DE MUC4 protein splice variant sv15.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ400852; CAC14137.1; -.
SQ SEQUENCE 1199 AA; 123130 MW; 135C88F6977EF11B CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1199;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASQGT 520
Db 783 STEASQGT 790

RESULT 72
Q9H486 PRELIMINARY; PRT; 1201 AA.
AC Q9H486;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE MUC4 protein splice variant sv16.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ400853; CAC14138.1; -.
SQ SEQUENCE 1201 AA; 123504 MW; 8C355BEAB245C5DC CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1201;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASQGT 520
Db 783 STEASQGT 790

RESULT 73
Q9H485 PRELIMINARY; PRT; 1214 AA.
AC Q9H485;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE MUC4 protein splice variant sv17.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";

```

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ400854; CAC14139.1; --
SQ SEQUENCE 1214 AA; 125245 MW; B89D6A84AD4A056 CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1214;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASGQT 520
DB 783 STEASGQT 790

RESULT 74

Q9H484 PRELIMINARY; PRT; 1215 AA.
AC Q9H484; (Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE MUC4 protein splice variant sv18.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancratic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ400855; CAC14140.1; --
SQ SEQUENCE 1215 AA; 125435 MW; 2B950CD46945D60D CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1215;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASGQT 520
DB 783 STEASGQT 790

RESULT 75

Q9H483 PRELIMINARY; PRT; 1256 AA.
AC Q9H483;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE MUC4 protein splice variant sv19.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancratic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ400856; CAC14141.1; --
SQ SEQUENCE 1256 AA; 130269 MW; 5BA45CE53651AADC CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1256;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASGQT 520
DB 783 STEASGQT 790

RESULT 76

Q8PKZ7 PRELIMINARY; PRT; 1290 AA.
AC Q8PKZ7; (Created)
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Hypothetical protein XAC2009.
GN XAC2009.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=32829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RC MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinaud F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Canarotte G., Cannavan F., Cardozo J., Chambergo F., Chapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.P.,
RA Locall E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sana J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.P., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.P.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing host specificities";
RL Nature 417:459-463(2002).
DR EMBL: AE011836; AAM36871.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1290 AA; 140955 MW; B9F16A206B21A01D CRC64;

Query Match 1.5%; Score 8; DB 16; Length 1290;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 VDAKIQQA 328
DB 313 VDAKIQQA 320

RESULT 77

Q9H482 PRELIMINARY; PRT; 1827 AA.
AC Q9H482;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE MUC4 protein splice variant sv20.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancratic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ400857; CAC14142.1; --
DR GO: 0016020; C-membrane; IEA.
DR GO: 0005215; P-transporter activity; IEA.
DR GO: 0007160; P-cell-matrix adhesion; IEA.


```
DR GO: 0006810; P:transport; IEA.
DR InterPro: IPR005533; AMOP.
DR InterPro: IPR003886; Nidogen ext.
DR InterPro: IPR005829; Sug transporter.
DR InterPro: IPR001846; VWF_D.
DR Pfam: PF03782; AMOP; 1.
DR Pfam: PF00094; vwd; 1.
DR SMART: SM00723; AMOP; 1.
DR SMART: SM00539; NIDO; 1.
DR SMART: SM00216; VWD; 1.
DR PROSITE: PS00217; SUGAR TRANSPORT 2; 1.
SQ SEQUENCE 1827 AA; 194100 MW; 63CE789B5C0F8741 CRC64;

Query Match      1.5%; Score 8; DB 4; Length 1827;
Best Local Similarity 100.0%; Pred.No.1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASGQT 520
Db 783 STEASGQT 790
|||||

RESULT 78
Q92JPS PRELIMINARY; PRT; 1902 AA.
ID AC Q92JPS
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cell surface antigen.
GN SCAR OR RC0019.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
DR ENBL: AE008570; AAL02557.1; -.
DR PIR: C97702; C97702.
DR GO: 0005509; F:calcium ion binding; IEA.
DR InterPro: IPR006315; Autotransporter.
DR InterPro: IPR005546; Autotransporter.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF03797; Autotransporter; 1.
DR TIGRFAMs: TIGR01414; autotrans_barl; 1.
DR PROSITE: PS00018; EF_HAND; 1.
KW Complete proteome.
SQ SEQUENCE 1902 AA; 212152 MW; F2D817457BF1912A CRC64;

Query Match      1.5%; Score 8; DB 16; Length 1902;
Best Local Similarity 100.0%; Pred.No.1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 KIKKLEKE 412
Db 356 KIKKLEKE 363
|||||

RESULT 79
Q9H4D8 PRELIMINARY; PRT; 2117 AA.
ID AC Q9H4D8
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mucin 4, variant V3.

Query Match      1.5%; Score 8; DB 4; Length 1827;
Best Local Similarity 100.0%; Pred.No.1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASGQT 520
Db 783 STEASGQT 790
|||||

RESULT 78
Q92JPS PRELIMINARY; PRT; 1902 AA.
ID AC Q92JPS
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cell surface antigen.
GN SCAR OR RC0019.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
DR ENBL: AE008570; AAL02557.1; -.
DR PIR: C97702; C97702.
DR GO: 0005509; F:calcium ion binding; IEA.
DR InterPro: IPR006315; Autotransporter.
DR InterPro: IPR005546; Autotransporter.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF03797; Autotransporter; 1.
DR TIGRFAMs: TIGR01414; autotrans_barl; 1.
DR PROSITE: PS00018; EF_HAND; 1.
KW Complete proteome.
SQ SEQUENCE 1902 AA; 212152 MW; F2D817457BF1912A CRC64;

Query Match      1.5%; Score 8; DB 4; Length 2117;
Best Local Similarity 100.0%; Pred.No.2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASGQT 520
Db 783 STEASGQT 790
|||||

RESULT 80
Q9H481 PRELIMINARY; PRT; 2167 AA.
ID AC Q9H481
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MUC4 protein splice variant sv21.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic tumor;
RX Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR ENBL: AJ400858; CAC14143.1; -.
DR GO: 0016020; C:membrane; IEA.
DR GO: 0005215; P:transporter activity; IEA.
DR GO: 0007160; P:cell-matrix adhesion; IEA.
DR GO: 0006810; P:transport; IEA.
DR InterPro: IPR005533; AMOP.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR006210; IEGF.
DR InterPro: IPR003886; Nidogen_ext.
DR InterPro: IPR005829; Sug transporter.
DR InterPro: IPR001846; VWF_D.
DR Pfam: PF03782; AMOP; 1.
DR Pfam: PF00008; EGF; 1.
DR Pfam: PF00094; vwd; 1.
```

DR SMART; SM00723; AMOP; 1.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00539; NIDO; 1.
DR SMART; SM00216; VMD; 1.
DR PROSITE; PS00022; EGF 1; 1.
DR PROSITE; PS00217; SUGAR TRANSPORT 2; 1.
DR PROSITE; PS00217; SUGAR TRANSPORT 2; 1.
SEQUENCE 2167 AA; 231235 MW; 3AF7A34178C22AF7 CRC64;
Query Match 1.5%; Score 8; DB 4; Length 2167;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 513 STEASGQT 520
DB 783 STEASGQT 790

RESULT 81
Q9NV09
ID Q9NV09 PRELIMINARY; PRT; 2169 AA.
AC Q9NV09
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Mucin 4 (MUC4).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
OX [1]
RN [1]
SEQUENCE FROM N.A.
MEDLINE=20381033; PubMed=10920259;
Choudhury A., Moniaux N., Wimpenny J.P., Hollingsworth M.A.,
Albert J.P., Batra S.K.;
"Human MUC4 mucin cDNA and its variants in pancreatic carcinoma.";
J. Biochem. 128:233-243(2000).
[2]
SEQUENCE FROM N.A.
Desseyn J.-L., Clavereau I., Laine A.;
"Cloning, chromosomal localization and characterization of the murine
mucin gene orthologous to human MUC4.";
Eur. J. Biochem. 269:0-0(2002).
[1]
DR ENBL; AJ276359; CAB81773.1; -.
DR ENBL; AF522055; AAM66747.1; -.
DR ENBL; AF522031; AAM66747.1; JOINED.
DR ENBL; AF522032; AAM66747.1; JOINED.
DR ENBL; AF522033; AAM66747.1; JOINED.
DR ENBL; AF522034; AAM66747.1; JOINED.
DR ENBL; AF522035; AAM66747.1; JOINED.
DR ENBL; AF522036; AAM66747.1; JOINED.
DR ENBL; AF522037; AAM66747.1; JOINED.
DR ENBL; AF522038; AAM66747.1; JOINED.
DR ENBL; AF522039; AAM66747.1; JOINED.
DR ENBL; AF522040; AAM66747.1; JOINED.
DR ENBL; AF522041; AAM66747.1; JOINED.
DR ENBL; AF522042; AAM66747.1; JOINED.
DR ENBL; AF522043; AAM66747.1; JOINED.
DR ENBL; AF522044; AAM66747.1; JOINED.
DR ENBL; AF522045; AAM66747.1; JOINED.
DR ENBL; AF522046; AAM66747.1; JOINED.
DR ENBL; AF522047; AAM66747.1; JOINED.
DR ENBL; AF522048; AAM66747.1; JOINED.
DR ENBL; AF522049; AAM66747.1; JOINED.
DR ENBL; AF522050; AAM66747.1; JOINED.
DR ENBL; AF522051; AAM66747.1; JOINED.
DR ENBL; AF522052; AAM66747.1; JOINED.
DR ENBL; AF522053; AAM66747.1; JOINED.
DR ENBL; AF522054; AAM66747.1; JOINED.
GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR005533; AMOP.

DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR005829; SUG_TRANSPORTER.
DR Pfam; PF03782; AMOP; 1.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00094; vmd; 1.
DR PROSITE; PS00022; EGF 1; 1.
DR PROSITE; PS00217; SUGAR TRANSPORT 2; 1.
DR PROSITE; PS00217; SUGAR TRANSPORT 2; 1.
SEQUENCE 2169 AA; 231470 MW; 5CEA1B16AF73E469 CRC64;
Query Match 1.5%; Score 8; DB 4; Length 2169;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 513 STEASGQT 520
DB 783 STEASGQT 790

RESULT 82
Q26676
ID Q26676 PRELIMINARY; PRT; 23 AA.
AC Q26676
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE SPOXTAL protein (Fragment).
GN SPOXTAL.
OS Tetrahydra aurantia (Orange pufferball sponge).
OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
OC Hadromerida; Tethyidae; Tethya.
OC NCBI_TaxID=34494;
OX [1]
RN [1]
SEQUENCE FROM N.A.
MEDLINE=95237605; PubMed=7721087;
Degan B.M., Degnan S.M., Giusti A., Morse D.E.;
"A box/hom homeobox gene in sponges.";
Gene 155:175-178(1995).
DR ENBL; X79265; CAA55852.1; -.
DR PIR; S45032; S45032.
FT NON_TER 1
FT NON_TER 23
FT NON_TER 23
SEQUENCE 23 AA; 2684 MW; 02682802A01FF3C2 CRC64;
Query Match 1.3%; Score 7; DB 5; Length 23;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 288 ELAERLK 294
DB 12 ELAERLK 18

RESULT 83
Q91K25
ID Q91K25 PRELIMINARY; PRT; 40 AA.
AC Q91K25
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=H20;
RA Gimenez-Barcons M., Franco S., Suarez Y., Fornes X., Ampurdanes S.,
Paig-Basagaiti F., Sanchez-Fueyo A., Barrera J.M., Llovet J.M.,
Brux J., Sanchez-Tapias J.M., Rodes J., Saiz J.C.;
"High amino acid variability within the NS5A gene of hepatitis C virus
(HCV) is associated with hepatocellular carcinoma in patients with
RT InterPro; IPR005533; AMOP.

[1]
RN SEQUENCE FROM N.A.
RP STRAIN-anti-D Recipient 68;
RX MEDLINE=98241727; PubMed=9573256;
RA McAllister J., Casino C., Davidson P., Power J., Lawlor E., Yap P.L.,
RA Simmonds P., Smith D.B.;
RT "Long-term evolution of the hypervariable region of hepatitis C virus
in a common-source-infected cohort";
RL J. Virol. 72:4893-4905(1998).
DR EMBL; AF056879; C: integral to membrane; IEA.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0019031; C: viral envelope; IEA.
DR GO; GO:0005198; F: structural molecule activity; IEA.
DR InterPro; IPR002519; HCV env.
DR Pfam; PF01539; HCV env; I.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane. 1
FT NON_TER 1
SQ SEQUENCE 47 AA; 4854 MW; 7C74EB8227AC7A23 CRC64;

Query Match 1.3%; Score 7; DB 12; Length 47;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GAQART 47
Db 25 GAQART 31

RESULT 86
Q8X2X0 PRELIMINARY; PRT; 60 AA.
AC Q8X2X0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein ECs4994.
GN ECs4994.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / RIMD 0509952;
RX MEDLINE=21156331; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:111-22(2001).
DR EMBL; AF002567; BAB38417.1; --
DR PIR; B91253; B91253.
KW Hypothetical protein.
SQ SEQUENCE 60 AA; 6634 MW; BAC2C287091F20C0 CRC64;

Query Match 1.3%; Score 7; DB 16; Length 60;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AQAPSP 500
Db 48 AQAPSP 54

RESULT 87
Q8YYQ3 PRELIMINARY; PRT; 62 AA.
ID Q8YYQ3
AC Q8YYQ3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)

RT HCV-1b related cirrhosis.";
RL Hepatology 0:0-0(2001).
DR EMBL; AF379247; AAK63464.1; --
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0019031; C: viral envelope; IEA.
DR GO; GO:0005198; F: structural molecule activity; IEA.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01560; HCV NS1; I.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 40
SQ SEQUENCE 40 AA; 4208 MW; A239A6F4CEB96829 CRC64;

Query Match 1.3%; Score 7; DB 12; Length 40;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GAQART 47
Db 18 GAQART 24

RESULT 84
Q8FHQ0 PRELIMINARY; PRT; 42 AA.
AC Q8FHQ0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN C1813.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raako D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Maynew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016760; AAN80277.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 42 AA; 4824 MW; BEE47C8780D1A75 CRC64;

Query Match 1.3%; Score 7; DB 16; Length 42;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 NKALLEM 430
Db 18 NKALLEM 24

RESULT 85
O72301 PRELIMINARY; PRT; 47 AA.
AC O72301;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;

01-MAR-2002 (TRENBLrel. 20, Last sequence update)
01-JUN-2003 (TRENBLrel. 24, Last annotation update)
Hypothetical protein ASL0793.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimori M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213 (2001).
DR EMBL; AP003583; BAB72750.1; --
DR PIR; AG1905; AG1905.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 62 AA; 7067 MW; 0229460C38C80422 CRC64;

Query Match 1.3%; Score 7; DB 16; Length 62;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 DLSAGGQ 477
|||||
DB 21 DLSAGGQ 27

RESULT 88
O72305 PRELIMINARY; PRT; 69 AA.
AC O72305;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=anti-D Recipient 68;
RX MEDLINE=98241727; PubMed=9573256;
RA McAllister J., Casino C., Davidson F., Power J., Lawlor E., Yap P.L.,
RA Simmonds P., Smith D.B.;
RT "Long-term evolution of the hypervariable region of hepatitis C virus
in a common-source-infected cohort.";
RL J. Virol. 72:4893-4905 (1998).
DR EMBL; AF056883; AAC17379.1; --
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 69
SQ SEQUENCE 69 AA; 7357 MW; F257BCCF7E4BFFAE CRC64;

Query Match 1.3%; Score 7; DB 12; Length 69;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GAQARTA 47
|||||

Db 12 GAQARTA 18

RESULT 89
Q8VCP1 PRELIMINARY; PRT; 70 AA.
AC Q8VCP1;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019472; AAH19472.1; --
KW Hypothetical protein.
SQ SEQUENCE 70 AA; 7944 MW; BCE7463EA6BA6CAB CRC64;

Query Match 1.3%; Score 7; DB 11; Length 70;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 NRSLSKEE 242
|||||
DB 43 NRSLSKEE 49

RESULT 90
Q9WLK7 PRELIMINARY; PRT; 74 AA.
AC Q9WLK7;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=anti-D Recipient 68;
RX MEDLINE=98241727; PubMed=9573256;
RA McAllister J., Casino C., Davidson F., Power J., Lawlor E., Yap P.L.,
RA Simmonds P., Smith D.B.;
RT "Long-term evolution of the hypervariable region of hepatitis C virus
in a common-source-infected cohort.";
RL J. Virol. 72:4893-4905 (1998).
DR EMBL; AF056885; AAC17381.1; --
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 74
SQ SEQUENCE 74 AA; 7777 MW; 89933164D292B86F CRC64;

Query Match 1.3%; Score 7; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GAQARTA 47
|||||


```
DE Hypothetical protein DR2007.
GN DR2007.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AB002038; AAF11565.1; -.
DR PIR; D75326; D75326.
DR TIGR; DR2007; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 79 AA; 8688 MW; 0E7D1F4162591C5B CRC64;

Query Match 1.3%; Score 7; DB 16; Length 79;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ELSRQLE 64
DB 12 ELSRQLE 18
|||||

RESULT 95
O63L13 PRELIMINARY; PRT; 81 AA.
AC O63L13;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Orf, partial conserved hypothetical protein.
GN Y121.G5 OR SF1580.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
DR EMBL; AS015179; AN43168.1; -.
DR InterPro; IPR002514; Transposase 8.
DR Pfam; PF01527; Transposase 8; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 81 AA; 8937 MW; AEEC05F025593C82 CRC64;

Query Match 1.3%; Score 7; DB 16; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 LKKEAVE 352
DB 56 LKKEAVE 62
|||||
```

```
RESULT 96
O87W81 PRELIMINARY; PRT; 82 AA.
AC O87W81;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN PSPT04673.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Pendorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collmer A.;
RT "Complete sequence of Pseudomonas syringae.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016872; AA058119.1; -.
DR TIGR; PSPT04673; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 82 AA; 8888 MW; 3826079B39FAD75C CRC64;

Query Match 1.3%; Score 7; DB 16; Length 82;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 384 TLSKSS 390
DB 35 TLSKSS 41
|||||

RESULT 97
O66133 PRELIMINARY; PRT; 84 AA.
AC O66133;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE BchZ (Fragment).
GN BCHZ.
OS Acidiphilium rubrum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Acetobacteraceae; Acidiphilium.
OX NCBI_TaxID=526;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98096867; PubMed=9435141;
RA Nagashima K.V., Matsuura K., Wakao N., Hiraishi A., Shimada K.;
RT "Nucleotide sequences of genes coding for photosynthetic reaction
RT centers and light-harvesting proteins of Acidiphilium rubrum and
RT related aerobic acidophilic bacteria.";
RL Plant Cell Physiol. 38:1249-1258(1997).
DR EMBL; AB005218; BAA25547.1; -.
FT NON_TER
SQ SEQUENCE 84 AA; 9230 MW; A404E530138BF61F CRC64;

Query Match 1.3%; Score 7; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 SPRVTEA 505
DB 66 SPRVTEA 72
|||||
```

```

RESULT 98
Q92JR7 PRELIMINARY; PRT; 84 AA.
AC Q92JR7
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein R03326.
GN R03326 OR SMC04092.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capella D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe P., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Gohl T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puebler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galbert F.;
RT Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591793; CAC47905.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 84 AA; 9237 MW; D4697D23X7259973 CRC64;

Query Match 1.3%; Score 7; DB 16; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 LKAEER 338
DB 42 LKAEER 48
|||||

RESULT 99
O66135 PRELIMINARY; PRT; 85 AA.
AC O66135
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE BchZ (Fragment).
GN BCHZ.
OS Acidiphilium angustum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Acetobacteraceae; Acidiphilium.
OX NCBI_TaxID=523;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98096867; PubMed=9435141;
RA Nagashima K.V., Matsura K., Wakao N., Hiraishi A., Shimada K.;
RT "Nucleotide sequences of genes coding for photosynthetic reaction
RT centers and light-harvesting proteins of Acidiphilium rubrum and
RT related aerobic acidophilic bacteria."
RL Plant Cell Physiol. 38:1249-1258(1997).
DR EMBL; AB005219; BAA25554.1; -.
FT NON TER 1
RP SEQUENCE 85 AA; 9343 MW; B84C69306C9EDC20 CRC64;

Query Match 1.3%; Score 7; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 SPRVTEA 505
DB 67 SPRVTEA 73
|||||

RESULT 100

```

```

O8L504 PRELIMINARY; PRT; 87 AA.
ID Q8L504
AC Q8L504
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative adenosine kinase (EC 2.7.1.20) (Fragment).
GN ADK.
OS Cicer arietinum (Chickpea) (Garbanzo).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Cicereae; Cicer.
OX NCBI_TaxID=3827;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Castellana; TISSUE=Epicotyl;
RA Dopico B., Labrador E.;
RT "An adenosine kinase is expressed in chickpea epicotyls."
RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ487468; CAD31841.1; -.
DR GO; GO:0004001; P:adenosine kinase activity; IEA.
DR GO; GO:0016301; P:kinase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR InterPro; IPR002173; PfKB.
DR Pfam; PF00294; pfkB; 1.
DR PROSITE; PS00584; PPKB_KINASES_2; 1.
KW Kinase; Transferase.
FT NON TER 1
RP SEQUENCE 87 AA; 9419 MW; 5703AA976B701A44 CRC64;

Query Match 1.3%; Score 7; DB 10; Length 87;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 SOLVQEK 208
DB 50 SOLVQEK 56
|||||

Search completed: June 7, 2004, 14:46:32
Job time : 50 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 12:38:45 ; Search time 139 Seconds
(without alignments)
1077.339 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 2702

Sequence: 1 KSSPCQBPAGPEGAQERDSQ.....APSTEASQGTQBPPTSARA 530

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2702	100.0	530	2	AAM49042 Human low
2	2702	100.0	530	4	AAB82804 Human low
3	2693	99.7	546	4	AAB82808 Human low
4	2693	99.7	639	6	ABR41210 Human DIT
5	2693	99.7	639	6	ABR41186 Human DIT
6	2465.5	91.2	557	2	AAM49039 Rabbit lo
7	2465.5	91.2	557	4	AAB82801 Rabbit lo
8	2338.5	86.5	1749	4	ABG00839 Novel hum
9	2324	86.0	510	4	ABR11764 Human LDL
10	2324	86.0	510	4	AAM79741 Human pro
11	1801	66.7	356	6	ABO07223 Human p53
12	1754.5	64.9	358	4	AAM78757 Human pro
13	1283.5	47.5	528	4	AAB93594 Human pro
14	1265.5	46.8	395	6	ABO07222 Human p53
15	1260.5	46.7	684	6	ABR43284 Human neu
16	1233	45.6	676	2	AAV17863 Neurite e
17	1038.5	38.4	386	2	AAM37883 BRCA1 mod
18	1038.5	38.4	386	2	AAV30151 Anitro aci
19	815	30.2	204	4	AAB63259 Human bre
20	810	30.0	475	4	Aao08301 Human pro
21	810	30.0	475	7	ADC33216 Human nov
22	659	24.4	416	7	ADC31613 Human nov
23	642.5	23.8	410	4	ABG02486 Novel hum
24	597.5	22.1	515	4	ABR61894 Drosophil
25	550	20.4	841	4	ABG02487 Novel hum

26	322	11.9	102	4	ABG02488 Novel hum
27	272	10.1	2274	4	ABR58657 Drosophil
28	265.5	9.9	954	4	AAU14615 Novel bon
29	265.5	9.9	1455	7	ADR80783 Microsate
30	257.5	9.5	860	7	ADC38517 Human AML
31	256	9.5	1752	2	AAU07031 Breast ca
32	256	9.5	2871	6	ABU07402 Protein d
33	256	9.5	2871	7	ADC35075 Human bre
34	256	9.5	2918	4	ABG27218 Novel hum
35	254.5	9.4	931	4	AAU79504 Human pro
36	254.5	9.4	990	4	AAU78520 Human pro
37	253.5	9.4	1881	5	ABP73809 Candida a
38	253	9.4	1206	7	ADD46507 Rat Prote
39	253	9.4	1206	7	ADD46507 Rat Prote
40	252.5	9.3	880	4	AAU96332 Putative
41	251.5	9.3	1690	4	ABR61144 Drosophil
42	251.5	9.3	1690	4	ABR61173 Drosophil
43	249	9.2	1392	2	AAU06999 Restin pr
44	248.5	9.2	786	2	AAU55937 Human SUL
45	248.5	9.2	1001	2	AAU55942 Human/Mur
46	248.5	9.2	1001	5	ABR97126 Novel hum
47	248.5	9.2	1001	7	ADE55356 Human pro
48	248.5	9.2	1005	6	ABR47509 Breast ca
49	248.5	9.2	1005	6	AAO30953 Human TAO
50	247	9.1	85	5	ABP42281 Human ova
51	247	9.1	1001	2	AAU48896 Rat TAO1
52	247	9.1	1001	7	ADE55355 Rat Prote
53	245	9.1	1427	2	AAU10534 Human 160
54	244.5	9.0	909	3	ABR53448 Human col
55	244.5	9.0	1240	4	AAU67538 Amino aci
56	244	9.0	1202	3	ABR08521 A murine
57	244	9.0	1212	6	ABP97029 Rat L-FIL
58	243.5	9.0	693	6	ADA14410 Mouse spe
59	242	9.0	1484	2	AAU89721 Canine ri
60	242	9.0	3259	7	ADE56037 Human pro
61	242	9.0	3259	7	ADE56033 Human pro
62	240	8.9	2633	4	ABG06505 Novel hum
63	239.5	8.9	1898	2	AAU30795 A human t
64	239.5	8.9	1898	7	ADD48869 Human pro
65	239	8.8	2192	2	AAU21732 LexA/NUMA
66	239	8.8	2272	2	AAU21731 GAL4/HA/N
67	239	8.8	2663	4	AAU39097 Human pol
68	239	8.8	2688	4	AAU40883 Human pol
69	238	8.8	892	7	ADB79862 Rat myosi
70	238	8.8	892	7	ADD47859 Rat prote
71	238	8.8	892	7	ADD47855 Rat prote
72	237.5	8.8	1213	4	AAU40016 Human pol
73	237.5	8.8	1213	6	ABP97031 Human L-F
74	236.5	8.8	1939	5	ABR77096 Human alp
75	236	8.7	1411	2	AAU02258 Nucleolar
76	235.5	8.7	1939	7	ADD47857 Human pro
77	235.5	8.7	1939	7	ADD47861 Human pro
78	235.5	8.7	1939	7	ADD47033 Human pro
79	234.5	8.7	650	7	ADB65482 Human pro
80	234.5	8.7	1235	5	ABR77435 Human tum
81	234.5	8.7	1235	7	ADD46509 Human pro
82	234.5	8.7	1235	7	ADD46509 Human pro
83	234.5	8.7	1646	4	ABR66368 Drosophil
84	234.5	8.7	1705	4	ABR71344 Drosophil
85	233	8.6	2115	3	AAU49937 Human NUM
86	233	8.6	2207	4	AAU32041 Novel hum
87	232	8.6	1935	7	ADD45207 Rat Prote
88	231.5	8.6	1886	2	AAU54241 Rattus no
89	231	8.5	1388	5	ABR57354 Mouse isc
90	230.5	8.5	860	7	ADC38519 MHC A pro
91	230.5	8.5	3187	7	ADE56031 Rat Prote
92	230.5	8.5	3187	7	ADE56035 Rat Prote
93	230.5	8.5	3899	6	ABR92048 Human cer
94	230.5	8.5	3907	6	ABR92047 Human cer
95	230.5	8.5	3911	7	ADB67140 Human A-K
96	230.5	8.5	3917	6	ABR92050 Human cer
97	230.5	8.5	3925	6	ABR92049 Human cer
98	230	8.5	1948	4	ABG21233 Novel hum

[illegible]

Db	361	KQETHLKOQALALYTEKPEFQNTLSKSSSEVFTTFKQEMEKMTKKIKKLEKETTMYSRW	420
Qy	421	ESSNKALLEMAEKTVDRELEGLQVKIORLEKLCRALQTERNDLNKRVQDLSAGGQSSL	480
Db	421	ESSNKALLEMAEKTVDRELEGLQVKIORLEKLCRALQTERNDLNKRVQDLSAGGQSSL	480
Qy	481	FTSGPERRPEGGGAQAPSSPRVTEAPCYFGAPSTEASGOTGQPEPTSARA	530
Db	481	FTSGPERRPEGGGAQAPSSPRVTEAPCYFGAPSTEASGOTGQPEPTSARA	530
RESULT 2			
ID	AAB82804	standard; protein; 530 AA.	
AC	AAB82804;		
XX			
DT	12-NOV-2001	(first entry)	
DE	Human low density lipoprotein binding protein 3 (LBP-3).		
KW	Low density lipoprotein binding protein 3; LBP-3; LDL; human;		
KM	atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.		
CS	Homo sapiens.		
XX			
PN	WO200164874-A2.		
XX			
PD	07-SEP-2001.		
XX			
PF	28-FEB-2001; 2001WO-US006356.		
XX			
PR	02-MAR-2000; 2000US-00517849.		
PR	14-JUL-2000; 2000US-00616289.		
XX			
PA	(BOST-) BOSTON HEART FOUND INC.		
XX			
PI	Lees AM, Lees RS, Law SW, Arjona AA;		
XX			
DR	WPI; 2001-565505/63.		
DR	N-PSDB; AAB26496.		
XX			
PT	New isolated low density lipoprotein binding polypeptide for treating,		
PT	diagnosing and/or identifying therapeutic agents for atherosclerosis.		
XX			
PS	Claim 13(h); Fig 9B; 143pp; English.		
XX			
CC	The present sequence is that of the N-terminal portion of novel human low		
CC	density lipoprotein binding protein 3 (LBP-3). The amino acid sequence is		
CC	deduced from an isolated partial cDNA clone (see AAB26494). A full-length		
CC	sequence is given in AAB82808. Human LBP-3 is an example of claimed LBP		
CC	polypeptides of the invention that are capable of binding to native and		
CC	methyalted low density lipoproteins. Also claimed are biologically active		
CC	fragments and analogues of LBPs, polynucleotides encoding LBPs, as well		
CC	as expression vectors, cells and methods of producing the LBPs. Methods		
CC	of determining if an animal is at risk for atherosclerosis, methods for		
CC	evaluating an agent for use in treating atherosclerosis, and methods for		
CC	treating a cell having an abnormality in structure or metabolism of LBP		
CC	are claimed. Pharmaceutical compositions comprising an LBP polypeptide or		
CC	nucleic acid, and vaccine compositions, are also claimed		
XX			
SQ	Sequence 530 AA;		
Query Match			
Best Local Similarity 100.0%; Score 2702; DB 4; Length 530;			
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	KSSPGQPEAGPEGAQBRPSQAAPVAEAGPGSSQAQPRKEGAQARTAQSGALRDVSEELS	60
Db	1	KSSPGQPEAGPEGAQBRPSQAAPVAEAGPGSSQAQPRKEGAQARTAQSGALRDVSEELS	60
Qy	61	ROLEDLSTYCVNNQGGPDGGAQGEPAEPEDAESKRTTYVARNGEPEPTPVVYGEKPS	120

Db	61	ROLEDLSTYCVNNQGGPDGGAQGEPAEPEDAESKRTTYVARNGEPEPTPVVYGEKPS	120
Qy	121	KGDPNTEBIROSDVGDHRRRPOKKAAGLGEKEITLLMQTLNTLSTPEEKLAALCKKY	180
Db	121	KGDPNTEBIROSDVGDHRRRPOKKAAGLGEKEITLLMQTLNTLSTPEEKLAALCKKY	180
Qy	181	AELLEHHNSOKMKLLQKQSQQLVQEKDHLRGEHSKAVLARSKLESLCRELQRNRSIK	240
Db	181	AELLEHHNSOKMKLLQKQSQQLVQEKDHLRGEHSKAVLARSKLESLCRELQRNRSIK	240
Qy	241	EEGVORAREEKEKKEVTSHFQVTLNDIQLQMEHNEHNSKLRQENMELARLKKLIQY	300
Db	241	EEGVORAREEKEKKEVTSHFQVTLNDIQLQMEHNEHNSKLRQENMELARLKKLIQY	300
Qy	301	ELREHIDKVFHKDLOQLVDKLLQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM	360
Db	301	ELREHIDKVFHKDLOQLVDKLLQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM	360
Qy	361	KQETHLKOQALALYTEKPEFQNTLSKSSSEVFTTFKQEMEKMTKKIKKLEKETTMYSRW	420
Db	361	KQETHLKOQALALYTEKPEFQNTLSKSSSEVFTTFKQEMEKMTKKIKKLEKETTMYSRW	420
Qy	421	ESSNKALLEMAEKTVDRELEGLQVKIORLEKLCRALQTERNDLNKRVQDLSAGGQSSL	480
Db	421	ESSNKALLEMAEKTVDRELEGLQVKIORLEKLCRALQTERNDLNKRVQDLSAGGQSSL	480
Qy	481	TDSGPERRPEGGGAQAPSSPRVTEAPCYFGAPSTEASGOTGQPEPTSARA	530
Db	481	TDSGPERRPEGGGAQAPSSPRVTEAPCYFGAPSTEASGOTGQPEPTSARA	530
RESULT 3			
ID	AAB82808	standard; protein; 546 AA.	
AC	AAB82808;		
XX			
DT	12-NOV-2001	(first entry)	
DE	Human low density lipoprotein binding protein 3 (LBP-3).		
KW	Low density lipoprotein binding protein 3; LBP-3; LDL; human;		
KM	atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200164874-A2.		
XX			
PD	07-SEP-2001.		
XX			
PF	28-FEB-2001; 2001WO-US006356.		
XX			
PR	02-MAR-2000; 2000US-00517849.		
PR	14-JUL-2000; 2000US-00616289.		
XX			
PA	(BOST-) BOSTON HEART FOUND INC.		
XX			
PI	Lees AM, Lees RS, Law SW, Arjona AA;		
XX			
DR	WPI; 2001-565505/63.		
DR	N-PSDB; AAB26501.		
XX			
PT	New isolated low density lipoprotein binding polypeptide for treating,		
PT	diagnosing and/or identifying therapeutic agents for atherosclerosis.		
XX			
PS	Claim 13(1); Fig 8A; 143pp; English.		
XX			
CC	The present sequence is that of novel human low density lipoprotein		
CC	binding protein 3 (LBP-3). The amino acid sequence was deduced from the		
CC	coding region of isolated genomic DNA (see AAB26501). It differs from the		
CC	sequence predicted from an LBP-3 cDNA clone (see AAB82804) by the		
CC	presence of an additional 16 amino acids at the N-terminus (the cDNA		
CC	clone is 5' truncated) and by having asparagine at amino acid position		

CC 130 rather than tyrosine. Human LBP-2 is an example of claimed LBP
 CC polypeptides of the invention that are capable of binding to native and
 CC methylated low density lipoproteins. Also claimed are biologically active
 CC fragments and analogues of these LBPs, polynucleotides encoding LBPs, as
 CC well as expression vectors, cells and methods of producing the LBPs. A
 CC polypeptide having the amino acid residues 96-110 of the present sequence
 CC is claimed (see A482820). Methods of determining if an animal is at risk
 CC for atherosclerosis, methods for evaluating a cell having an abnormality in
 CC atherosclerosis, and methods for treating a cell having an abnormality in
 CC structure or metabolism of LBP are claimed. Pharmaceutical compositions
 CC comprising an LBP polypeptide or nucleic acid, and vaccine compositions,
 CC are also claimed
 XX
 SQ Sequence 546 AA;
 Query Match 99.7%; Score 2693; DB 4; Length 546;
 Best Local Similarity 99.8%; Pred. No. 1.9e-159;
 Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KSPGQPEAGPEGAQRPSPQAAPVAEAGPGSSQAQPRKPEGAQARTAQSGALADVSEELS 60
 Db 17 KSPGQPEAGPEGAQRPSPQAAPVAEAGPGSSQAQPRKPEGAQARTAQSGALADVSEELS 76
 QY 61 RQLEDILSTVCVNNQGGPDEGAQGEPAEPDAEKSRTYVARNGPEPTPVVYGEKPS 120
 Db 77 RQLEDILSTVCVNNQGGPDEGAQGEPAEPDAEKSRTYVARNGPEPTPVVYGEKPS 136
 QY 121 KGPNTPEIRQSDVEGDRHRRPOEKKAKGLGKITLLMOTLNTLSTPEEKLAALCKKY 180
 Db 137 KGPNTPEIRQSDVEGDRHRRPOEKKAKGLGKITLLMOTLNTLSTPEEKLAALCKKY 196
 QY 181 AELLLEHNSOKMKLQKQSOVLQVQKHLCRGEHSHKAVLARSLSLCLRELQHRNSLK 240
 Db 197 AELLLEHNSOKMKLQKQSOVLQVQKHLCRGEHSHKAVLARSLSLCLRELQHRNSLK 256
 QY 241 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNERNSKLRQENNELAERLKLIEQY 300
 Db 257 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNERNSKLRQENNELAERLKLIEQY 316
 QY 301 ELAEEHIDKVKHKLQOQIYDVKLQQAQEMKAEERHQREKDFLLKAVESQRCMLM 360
 Db 317 ELAEEHIDKVKHKLQOQIYDVKLQQAQEMKAEERHQREKDFLLKAVESQRCMLM 376
 QY 361 KQOETHLKQALAYTEKFEFQNTLSKSSVFVTFQEMKMTKKIKCLEKETTMYRSRW 420
 Db 377 KQOETHLKQALAYTEKFEFQNTLSKSSVFVTFQEMKMTKKIKCLEKETTMYRSRW 436
 QY 421 ESSNKALLEVAEKTVKDEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSLL 480
 Db 437 ESSNKALLEVAEKTVKDEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSLL 496
 QY 481 TDGSPRRRPGGAQAPSPSRVTEAPCYPGAPSTESAGTGPQEPPTSARA 530
 Db 497 TDGSPRRRPGGAQAPSPSRVTEAPCYPGAPSTESAGTGPQEPPTSARA 546
 RESULT 4
 ABR41210
 ID ABR41210 standard; protein; 639 AA.
 XX
 AC ABR41210;
 XX
 DT 02-JUN-2003 (first entry)
 XX
 DE Human DITHP extracellular signalling protein.
 XX
 KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
 KW cancer; cell proliferative disorder; autoimmune disorder;
 KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
 KW neurological disorder; gastrointestinal disorder; transport disorder;
 KW connective tissue disorder; drug screening; proteomic analysis;
 KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
 KW disease model; toxicological testing; transcript imaging;

extracellular signalling.
 Homo sapiens.
 WO200297031-A2.
 05-DEC-2002.
 27-MAR-2002; 2002WO-US010056.
 28-MAR-2001; 2001US-0279619P.
 29-MAR-2001; 2001US-0280067P.
 16-MAY-2001; 2001US-0280068P.
 17-MAY-2001; 2001US-0291280P.
 17-MAY-2001; 2001US-0291849P.
 19-JUN-2001; 2001US-0299428P.
 20-JUN-2001; 2001US-0299776P.
 20-JUN-2001; 2001US-0300001P.
 (INCY-) INCYTE GENOMICS INC.
 Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
 Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;
 Daughtery SC, Dam TC, Liu TP, Nguyen DA, Kleefeld Y, Gerstin EH;
 Peralta CH, David KH, Lewis SA, Chen AJ, Panzer SR, Harris B;
 Flores V, Marwaha R, Lo A, Ian RY, Urashka ME;
 WPI; 2003-129518/12.
 N-PSDB; ACC46153.
 Novel human diagnostic and therapeutic polypeptide useful for identifying
 test compound which specifically binds to a polypeptide encoded by human
 diagnostic and therapeutic polynucleotide, and to induce antibodies.
 Claim 27; SEQ ID NO 745; 591pp; English.
 The invention relates to novel human diagnostic and therapeutic
 polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
 proteins (DITHP: ABR41136-ABR41812). The invention also relates to
 polynucleotide sequences at least 90% identical to the dithp cDNA
 sequences of the invention; recombinant vectors, host cells and
 transgenic organisms comprising a dithp nucleic acid sequence; the
 recombinant production of DITHP proteins; antibodies specific for DITHP
 proteins; microarrays comprising dithp nucleic acid sequences; methods of
 detecting dithp nucleotide and protein sequences; methods of screening
 for compounds which specifically bind a DITHP protein; and methods of
 assessing the toxicity of test compounds using a dithp hybridisation
 probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
 diagnosis of a wide variety of conditions including cancer and other cell
 proliferative disorders; autoimmune or inflammatory disorders; bacterial,
 viral, fungal or parasitic infections; hormonal disorders; metabolic
 disorders; neurological disorders; gastrointestinal disorders; transport
 disorders; and connective tissue disorders. They may also be used to
 screen for modulators of protein activity or gene expression. DITHP
 proteins can additionally be used in analysis of the proteome of a tissue
 or cell type and to induce antibodies. The dithp nucleic acids are
 additionally useful in somatic or germline gene therapy of the disorders
 mentioned above, as a source of antisense sequences, as a source of
 probes and primers, in genotyping and identification of individuals, in
 the generation of transgenic animal models of human disease or knock in
 humanised animals, in toxicological testing, and in transcript imaging.
 The present sequence represents a DITHP protein which has extracellular
 signalling activity. Note: The sequence data for this patent did not form
 part of the printed specification, but was obtained in electronic format
 directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 639 AA;
 Query Match 99.7%; Score 2693; DB 6; Length 639;
 Best Local Similarity 99.8%; Pred. No. 2.2e-159;
 Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKPEGAQARTAQSGALRDVSEELS 60
DB 110 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKPEGAQARTAQSGALRDVSEELS 169
QY 61 RQLEDILSTYCVNNQGGGEGAGGAEPAEDAEKSTTYVARNGEPEPTPVVYGEKPS 120
DB 170 RQLEDILSTYCVNNQGGGEGAGGAEPAEDAEKSTTYVARNGEPEPTPVVYGEKPS 229
QY 121 KGDPTNTEIROSDEVDGRDHRPQEKKAAGLGEITLLMOTLNTLSTPEEKLAALCKKY 180
DB 230 KGDPTNTEIROSDEVDGRDHRPQEKKAAGLGEITLLMOTLNTLSTPEEKLAALCKKY 289
QY 181 AELLEHRNSQOMKLLQKQSQVLQVQKHRLRGHSHKAVLARSKLESICRELQHRNLSLK 240
DB 290 AELLEHRNSQOMKLLQKQSQVLQVQKHRLRGHSHKAVLARSKLESICRELQHRNLSLK 349
QY 241 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNRSKLQENMELARLKLIEQY 300
DB 350 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNRSKLQENMELARLKLIEQY 409
QY 301 ELREBHIDKVPKHKDLOQOVLDAKLOQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 360
DB 410 ELREBHIDKVPKHKDLOQOVLDAKLOQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 469
QY 361 KQOETHLQKQALYKTEKPEEFONTLSKSSEVFTTFQEMEKMTKKIKLEKETMYRSRW 420
DB 470 KQOETHLQKQALYKTEKPEEFONTLSKSSEVFTTFQEMEKMTKKIKLEKETMYRSRW 529
QY 421 ESNVALLMABEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLS 480
DB 530 ESNVALLMABEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLS 589
QY 481 TDSGPERPEGCAQAPSPRVTEAPCPYGPAPSTASGOTGPOEPTSARA 530
DB 590 TDSGPERPEGCAQAPSPRVTEAPCPYGPAPSTASGOTGPOEPTSARA 639
RESULT 5
ID ABR41186
AC ABR41186 standard; protein; 639 AA.
AC ABR41186;
DT 02-JUN-2003 (first entry)
DE Human DITHP extracellular signalling protein.
KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW extracellular signalling.
OS Homo sapiens.
XX
XX W0200297031-A2.
XX
XX 05-DEC-2002.
XX
XX 27-MAR-2002; 2002WO-US010056.
XX
XX 28-MAR-2001; 2001US-0279619P.
XX 29-MAR-2001; 2001US-0280067P.
XX 29-MAR-2001; 2001US-0280068P.
XX 16-MAY-2001; 2001US-0291280P.
XX 17-MAY-2001; 2001US-0291829P.
XX 17-MAY-2001; 2001US-0291849P.
XX 19-JUN-2001; 2001US-0299428P.
XX 20-JUN-2001; 2001US-0299776P.
XX 20-JUN-2001; 2001US-0300001P.

(INCY-) INCYTE GENOMICS INC.
Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;
Daugherty SC, Dam TC, Liu TF, Nguyen DA, Klesfeld Y, Gerstein EH;
Peralta CH, David MH, Lewis SA, Chen AH, Panzer SR, Harris B,
Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
WPI; 2003-129518/12.
DR N-PSDB; ACC46130.
XX
XX Novel human diagnostic and therapeutic polypeptide useful for identifying
XX test compound which specifically binds to a polypeptide encoded by human
XX diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX
XX Claim 27; SEQ ID NO 721; 591bp; English.
XX
XX The invention relates to novel human diagnostic and therapeutic
XX polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
XX proteins (DITHP; ABR41136-ABR41812). The invention also relates to
XX polynucleotide sequences at least 90% identical to the dithp cDNA
XX sequences of the invention; recombinant vectors, host cells and
XX transgenic organisms comprising a dithp nucleic acid sequence; the
XX recombinant production of DITHP proteins; antibodies specific for DITHP
XX proteins; microarrays comprising dithp nucleic acid sequences; methods of
XX detecting dithp nucleotide and protein sequences; methods of screening
XX for compounds which specifically bind a DITHP protein; and methods of
XX assessing the toxicity of test compounds using a dithp hybridisation
XX probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
XX diagnosis of a wide variety of conditions including cancer and other cell
XX proliferative disorders; autoimmune or inflammatory disorders; bacterial,
XX viral, fungal or parasitic infections; hormonal disorders; metabolic
XX disorders; neurological disorders; gastrointestinal disorders; transport
XX disorders; and connective tissue disorders. They may also be used to
XX screen for modulators of protein activity or gene expression. DITHP
XX proteins can additionally be used in analysis of the proteome of a tissue
XX or cell type and to induce antibodies. The dithp nucleic acids are
XX additionally useful in somatic or germline gene therapy of the disorders
XX mentioned above, as a source of antisense sequences, as a source of
XX probes and primers, in genotyping and identification of individuals, in
XX the generation of transgenic animal models of human disease or knock in
XX humanised animals, in toxicological testing, and in transcript imaging.
XX The present sequence represents a DITHP protein which has extracellular
XX signalling activity. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 639 AA;
Query Match 99.7%; Score 2693; DB 6; Length 639;
Best Local Similarity 99.8%; Pred. No. 2.2e-159;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKPEGAQARTAQSGALRDVSEELS 60
DB 110 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKPEGAQARTAQSGALRDVSEELS 169
QY 61 RQLEDILSTYCVNNQGGGEGAGGAEPAEDAEKSTTYVARNGEPEPTPVVYGEKPS 120
DB 170 RQLEDILSTYCVNNQGGGEGAGGAEPAEDAEKSTTYVARNGEPEPTPVVYGEKPS 229
QY 121 KGDPTNTEIROSDEVDGRDHRPQEKKAAGLGEITLLMOTLNTLSTPEEKLAALCKKY 180
DB 230 KGDPTNTEIROSDEVDGRDHRPQEKKAAGLGEITLLMOTLNTLSTPEEKLAALCKKY 289
QY 181 AELLEHRNSQOMKLLQKQSQVLQVQKHRLRGHSHKAVLARSKLESICRELQHRNLSLK 240
DB 290 AELLEHRNSQOMKLLQKQSQVLQVQKHRLRGHSHKAVLARSKLESICRELQHRNLSLK 349
QY 241 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNRSKLQENMELARLKLIEQY 300
DB 350 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNRSKLQENMELARLKLIEQY 409

QY	301	ELRREHDKVFKHDLQQLVDALQQAQENLKEABERHQREKDFLLKEAVESQRMCELM	360
Db	410	ELRREHDKVFKHDLQQLVDALQQAQENLKEABERHQREKDFLLKEAVESQRMCELM	469
QY	361	KQOETHLKOALALYTEKEPEBQNTLSKSSVFITFQEMSEKMTKKIKLLEKETTMYRSRW	420
Db	470	KQOETHLKOALALYTEKEPEBQNTLSKSSVFITFQEMSEKMTKKIKLLEKETTMYRSRW	529
QY	421	ESSNKALLEMAEKTVRDKEGLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQQSLS	480
Db	530	ESSNKALLEMAEKTVRDKEGLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQQSLS	589
QY	481	TDSGPERPEGPGAQAPSPRVTEAPCYPGAPSTEASGQTGPQEP7SARA	530
Db	590	TDSGPERPEGPGAQAPSPRVTEAPCYPGAPSTEASGQTGPQEP7SARA	639
RESULT 6			
AAW49039	ID AAW49039 standard; protein; 557 AA.		
ID	XX AAW49039;		
AC	XX 09-NOV-1998 (first entry)		
DT	XX Rabbit low density lipoprotein binding protein LBP-3.		
DE	XX Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;		
KW	XX receptor; rabbit; atherosclerosis; diagnosis; therapy; vaccine.		
KW	XX Oryctolagus cuniculus.		
OS	XX		
Key	XX Location/Qualifiers		
PE	XX Peptide 96..110		
FT	XX /note= "Claim 2"		
FT	XX WO9823282-A1.		
PN	XX 04-JUN-1998.		
PD	XX 26-NOV-1997; 97WO-US021857.		
PF	XX 27-NOV-1996; 96US-0031930P.		
PR	XX 03-JUN-1997; 97US-0048547P.		
XX	(BOST-) BOSTON HEART FOUND INC.		
PA	XX Lees AM, Lees RS, Law SW, Arjona AA;		
PI	XX WPI; 1998-322455/28.		
DR	XX N-PSDB; AAV32836.		
XX	XX		
PT	XX Nucleic acid encoding low density lipoprotein binding proteins and		
PT	XX related vectors - transformed cells, proteins, and modulators of binding,		
PT	XX useful for treatment and diagnosis of atherosclerosis and for identifying		
PT	XX subjects at risk.		
XX	XX		
XX	XX Claim 1; Fig 5; 47pp; English.		
XX	XX This polypeptide comprises novel rabbit low density lipoprotein (LDL)		
CC	XX binding protein Lbp-3 that is capable of binding both native and methyl		
CC	XX LDL. Its amino acid sequence was deduced from rabbit abdominal aorta cDNA		
CC	XX (see AAV32836). cDNA clones (see AAV32834-39) antecoded rabbit and human		
CC	XX LBP3 (see AAW49037-42) are claimed. An abnormality in an aspect of LBP		
CC	XX metabolism or structure is diagnostic of a risk for atherosclerosis. The		
CC	XX invention provides methods for determining if an animal is at risk for		
CC	XX atherosclerosis (e.g. for prenatal screening); methods for treating		
CC	XX atherosclerosis (including gene therapy) using e.g. LBP polypeptides to		
CC	XX bind LDL and thereby prevent formation of atherosclerotic plaques; and		
CC	XX methods for treating a cell having an abnormality in LBP structure or		
CC	XX metabolism. Pharmaceutical and vaccine compositions are also provided, as		
CC	XX well as recombinant vectors and host cells used to produce recombinant		

Db 228 RQLEDILSTYCVNNOGGPGEAGQAPPAEDAEKSRRTTVARNGEPEPTPVVNGEKPS 287
QY 121 KGDPTNTEIRQSDVEGDRHRRPQEKKAAGL-----GKSTLLMOTLNTLSTPE 170
Db 288 KGDPTNTEIRQSDVEGDRHRRPQEKKAAGLGEQRAALCEAGKEITLLMOTLNTLSTPE 347
QY 171 EKLAALCKKYAHLLEHRSQKMKLLQKQSQLVQEKDHLRGEHSAVLARSKLESICR 230
Db 348 EKLAALCKKYAHLLEHRSQKMKLLQKQSQLVQEKDHLRGEHSAVLARSKLESICR 407
QY 231 ELQRHNRSLK-----SEGQVBARBEERKEKVTSHFQVTLNDIOL 270
Db 408 ELQRHNRSLKGDPCVAVDDGSRADPFAQEGVQARBEERKEKVTSHFQVTLNDIOL 467
QY 271 QMEQHNRSLKQENMELARLKLLEQVELBEHIDKFKHKLQOOLVDAKLOOAOE 330
Db 468 QMEQHNRSLKQENMELARLKLLEQVELBEHIDKFKHKLQOOLVDAKLOOAOE 527
QY 331 MKAEERHQRKDFLLKEAVESQRMCELMKQOETHLQOOLALYTEKFEFQNTLSKSE 390
Db 528 MKAEERHQRKDFLLKEAVESQRMCELMKQOETHLQOOLALYTEKFEFQNTLSKSE 587
QY 391 VFTTFQKQEKMTKKIKLEKETTMYRSRWESENKALLEMAEKTVDKLEGLQVKIOR 450
Db 588 VFTTFQKQEKMTKKIKLEKETTMYRSRWESENKALLEMAEKTVDKLEGLQVKIOR 647
QY 451 LEKLCRALQTERNDLNRVQDL-----SAGGCSLTDSGPERRPEGCAQAP 497
Db 648 LEKLCRALQTERNDLNRVQDL-----SAGGCSLTDSGPERRPEGCAQAP 707
QY 498 SS-----PRVTEAPCYPGAPSTASQGTGPQP 525
Db 708 TSARKRPNR---KPLLPG---DGEGLRPSQRP 733

RESULT 9
ABBI1764
ID ABBI1764 standard; peptide; 510 AA.
XX AC ABBI1764;
XX DT 11-JAN-2002 (first entry)
XX DE Human LDL binding protein homologue, SEQ ID NO:2134.
XX KW Human; cytokine; cell proliferation; tissue growth; immunomodulator; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antiulcer.
XX OS Homo sapiens.
XX PN WO200157188-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-US003900.
XX PR 03-FEB-2000; 2000US-00496914.
XX PR 27-APR-2000; 2000US-00560875.
XX XX (HYSE-) HYSEQ INC.
XX PA Tang YT, Liu C, Drmanac RT;
XX PI

XX WPI, 2001-457740/49.
DR N-PsDB; ABA09008.
XX Human proteins and DNA encoding sequences useful for preventing, treating
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
PT and cancer.
XX Claim 20; Page 240-241; 1963pp; English.

XX Sequences ABBI0991-ABBI2330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g. asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention

XX SQ Sequence 510 AA;
Query Match 86.0%; Score 2324; DB 4; Length 510;
Best Local Similarity 99.8%; Pred. No. 1.6e-136;
Matches 459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAPAVEAGFGSSQAPRKPEGAQARTAQSGALDVSSELS 60
Db 28 KSSPGQPEAGPEGAQERPSQAPAVEAGFGSSQAPRKPEGAQARTAQSGALDVSSELS 87
QY 61 RQLEDILSTYCVNNOGGPGEAGQAPPAEDAEKSRRTTVARNGEPEPTPVVNGEKPS 120
Db 88 RQLEDILSTYCVNNOGGPGEAGQAPPAEDAEKSRRTTVARNGEPEPTPVVNGEKPS 147
QY 121 KGDPTNTEIRQSDVEGDRHRRPQEKKAAGLGEITLLMOTLNTLSTPEKLAALCKKY 180
Db 148 KGDPTNTEIRQSDVEGDRHRRPQEKKAAGLGEITLLMOTLNTLSTPEKLAALCKKY 207
QY 181 AEILLEHRSQKMKLLQKQSQLVQEKDHLRGEHSAVLARSKLESICRQLRHRSLK 240
Db 208 AEILLEHRSQKMKLLQKQSQLVQEKDHLRGEHSAVLARSKLESICRQLRHRSLK 267
QY 241 EGVQVBARBEERKEKVTSHFQVTLNDIOLQMEQHNRSLKQENMELARLKLLEQVEL 300
Db 268 EGVQVBARBEERKEKVTSHFQVTLNDIOLQMEQHNRSLKQENMELARLKLLEQVEL 327

QY 301 ELREHIDKVFHKDLQQLVDKALQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
 DB 328 ELREHIDKVFHKDLQQLVDKALQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 387
 QY 361 KQETHLKKQALALYTEKPEFQNTLSKSSSVFTTFKQEMERQTKIKKLEKETTMYRSRW 420
 DB 388 KQETHLKKQALALYTEKPEFQNTLSKSSSVFTTFKQEMERQTKIKKLEKETTMYRSRW 447
 QY 421 ESSNKALLEMAEBKTVRDKEGLQVKIQRLEKLCRALQT 460
 DB 448 ESSNKALLEMAEBKTVRDKEGLQVKIQRLEKLCRALQT 487
 RESULT 10
 ID AAM79741
 ID AAM79741 standard; protein; 510 AA.
 AC AAM79741;
 DT 06-NOV-2001 (first entry)
 DE Human protein SEQ ID NO 3387.
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX Homo sapiens.
 OS
 FN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US004098.
 PR 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 DR WPI: 2001-476283/51.
 DR N-PSDB; AAK52874.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX
 PS Claim 20; Page 319; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication

XX SQ Sequence 510 AA;
 Query Match 86.0%; Score 2324; DB 4; Length 510;
 Best local Similarity 99.8%; Pred. No. 1.66-136;
 Matches 459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKPEGAQARTAQSGALRDVSELS 60
 DB 28 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKPEGAQARTAQSGALRDVSELS 87
 QY 61 RQLEDILSTYCVNNQGGPGEDGCAQGEPAPEPAEKSRITYVARGPEPTPVVYGEKPS 120
 DB 88 RQLEDILSTYCVNNQGGPGEDGCAQGEPAPEPAEKSRITYVARGPEPTPVVYGEKPS 147
 QY 121 KQDPNTEIRQSDDEVGDHRRPOEKKKAKGLGKEITLLMQTLNTLSTPEKLAALCKKY 180
 DB 148 KQDPNTEIRQSDDEVGDHRRPOEKKKAKGLGKEITLLMQTLNTLSTPEKLAALCKKY 207
 QY 181 AELLEHRNSQKMKLLQKQKQSVLQVQKHLSREHSAVLARSKLSLCSLCELOHNRSLK 240
 DB 208 AELLEHRNSQKMKLLQKQKQSVLQVQKHLSREHSAVLARSKLSLCSLCELOHNRSLK 267
 QY 241 EKGQVQARREERKEKVTSHFQVTLNDIQLQMHNRNSKLRQENNELAERLKLIEQY 300
 DB 268 EKGQVQARREERKEKVTSHFQVTLNDIQLQMHNRNSKLRQENNELAERLKLIEQY 327
 QY 301 ELREHIDKVFHKDLQQLVDKALQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
 DB 328 ELREHIDKVFHKDLQQLVDKALQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 387
 QY 361 KQETHLKKQALALYTEKPEFQNTLSKSSSVFTTFKQEMERQTKIKKLEKETTMYRSRW 420
 DB 388 KQETHLKKQALALYTEKPEFQNTLSKSSSVFTTFKQEMERQTKIKKLEKETTMYRSRW 447
 QY 421 ESSNKALLEMAEBKTVRDKEGLQVKIQRLEKLCRALQT 460
 DB 448 ESSNKALLEMAEBKTVRDKEGLQVKIQRLEKLCRALQT 487
 RESULT 11
 ID ABO07223
 ID ABO07223 standard; protein; 356 AA.
 AC ABO07223;
 XX
 DT 13-AUG-2003 (first entry)
 DE Human p53 modifying protein, SEQ ID 183.
 KW Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;
 KW antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
 KW lung cancer; ovarian cancer; angiogenesis; cell cycle;
 KW apoptotic disorder; cell proliferation disorder.
 XX Homo sapiens.
 OS
 FN WO200299122-A1.
 XX
 PD 12-DEC-2002.
 XX
 PF 03-JUN-2002; 2002WO-US017382.
 XX
 PR 05-JUN-2001; 2001US-0296076P.
 PR 10-OCT-2001; 2001US-0328605P.
 PR 15-FEB-2002; 2002US-0357253P.
 XX (EXEL-) EXELIXIS INC.
 PA
 XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
 PI WPI: 2003-156859/15.
 DR N-PSDB; ACTD1344B.

XX Identifying modulators of the p53 pathway for use in treating apoptotic
PT or cell proliferation disorders, comprises screening for agents that
PT modulate activity of a human ortholog of genes that modify the p53
PT pathway in Drosophila.

XX Example 2; Page 533-534; 678pp; English.

XX The invention relates to identifying (M1) a candidate p53 pathway
XX modulating agent, by contacting an assay system comprising a purified HM
CC polypeptide (human orthologue of genes that modify the p53 pathway in
CC Drosophila) or nucleic acid with a test agent under conditions, where but
CC for the presence of the test agent, the system provides a reference
CC activity, and detecting a test agent-biased activity of the assay system.
CC Also included are modulating (M2) a p53 pathway of a cell (comprising
CC contacting a cell defective in p53 function with a candidate modulator
CC that specifically binds to a HM polypeptide comprising an HM amino acid
CC sequence, where p53 function is restored), modulating (M3) a p53 pathway
CC in a mammalian cell (comprising contacting the cell with an agent that
CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
CC a disease in a patient (comprising: (a) obtaining a biological sample
CC from the patient; (b) contacting the sample with a probe for HM
CC expression; (c) comparing the results with a control; and (d) determining
CC whether the comparison indicates a likelihood disease). (M1) is useful
CC for identifying modulators of the p53 pathway. A probe for HM expression
CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
CC in a patient, where the cancer has greater than 25 % expression level.
CC Modulators identified by (M1) are useful in a variety of diagnostic and
CC therapeutic applications, where disease or disorder prognosis is related
CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
CC proliferation disorders (e.g. cancer). Another two new methods (M2 and
CC M3) are useful for modulating the p53 pathway of a cell, thus restoring
CC the p53 function of the cell, so that the cell undergoes normal
CC proliferation or progression through the cell cycle. (M2) and (M3) are
CC also useful for treating defects in the p53 pathway such as angiogenic,
CC apoptotic or cell proliferation disorders. The present sequence
CC represents a human p53 pathway modifying protein

XX Sequence 356 AA;

Query Match 66.7%; Score 1801; DB 6; Length 356;
Best Local Similarity 100.0%; Pred. No. 3.7e-104;
Matches 356; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 ALCKYAELEHRSNOKMLOKQSQOLVQKDLRGEHSAVLARSKLSLCRELQR 234
Db 1 ALCKYAELEHRSNOKMLOKQSQOLVQKDLRGEHSAVLARSKLSLCRELQR 60

QY 235 HNSLKEGVQARAREEERKEVTSFQVTLNDIQLOMEQHNRNSKLRQENNELAERLKK 294
Db 61 HNSLKEGVQARAREEERKEVTSFQVTLNDIQLOMEQHNRNSKLRQENNELAERLKK 120

QY 295 KLEIQVELREEHIDKVFVKHDLQOOLVDAKLQQAQEMLKEAERHQRKDFLLKEAVESQ 354
Db 121 KLEIQVELREEHIDKVFVKHDLQOOLVDAKLQQAQEMLKEAERHQRKDFLLKEAVESQ 180

QY 355 RMCLEMKQOETHLKOQALALYTKFEEFQNTLSKSEVFTTFQKQEMKMTKIKLEKETTT 414
Db 181 RMCLEMKQOETHLKOQALALYTKFEEFQNTLSKSEVFTTFQKQEMKMTKIKLEKETTT 240

QY 415 MYRSRWSSNNKALLEMAEAKTVRDKLEGLQVKIQLEKLCRALQTERNDLNKRVQDLISA 474
Db 241 MYRSRWSSNNKALLEMAEAKTVRDKLEGLQVKIQLEKLCRALQTERNDLNKRVQDLISA 300

QY 475 GGQGSLLTDSGPERPBGPGQAQAPSSPRVTEAPCYPCGAPSTEASGGTGPOEPTTSARA 530
Db 301 GGQGSLLTDSGPERPBGPGQAQAPSSPRVTEAPCYPCGAPSTEASGGTGPOEPTTSARA 356

RESULT 12
AAM78757
ID AAM78757
XX

AC AAM78757;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1419.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX Homo sapiens.
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US004098.
XX 03-FEB-2000; 2000US-00496914.
XX 27-APR-2000; 2000US-00560875.
XX 20-JUN-2000; 2000US-00598075.
XX 19-JUL-2000; 2000US-00620325.
XX 01-SEP-2000; 2000US-00654936.
XX 15-SEP-2000; 2000US-00663561.
XX 20-OCT-2000; 2000US-00693325.
XX 30-NOV-2000; 2000US-00728422.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
XX N-PSDB; AAK51890.
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX Claim 20; Page 3681-3682; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication

XX Sequence 358 AA;

Query Match 64.9%; Score 1754.5; DB 4; Length 358;
Best Local Similarity 98.3%; Pred. No. 2.9e-101;
Matches 349; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 177 CKKY-AELLEHRSNOKMLOKQSQOLVQKDLRGEHSAVLARSKLSLCRELQR 235
Db 4 CFPYPCPLEHRSNOKMLOKQSQOLVQKDLRGEHSAVLARSKLSLCRELQR 63

QY 236 NRSLEKEGVQARAREEERKEVTSFQVTLNDIQLOMEQHNRNSKLRQENNELAERLKK 295
Db 64 NRSLEKEGVQARAREEERKEVTSFQVTLNDIQLOMEQHNRNSKLRQENNELAERLKK 123

QY 296 LIEQVELREEHIDKVFVKHDLQOOLVDAKLQQAQEMLKEAERHQRKDFLLKEAVESQ 355
Db 124 LIEQVELREEHIDKVFVKHDLQOOLVDAKLQQAQEMLKEAERHQRKDFLLKEAVESQ 183

QY 356 MCELKMQQETHLKKQALALYTEKFEFFQNTLSKSGSEVFTTFQEMKMKTKKIKKLEKETTMM 415
Db 184 MCELKMQQETHLKKQALALYTEKFEFFQNTLSKSGSEVFTTFQEMKMKTKKIKKLEKETTMM 243
QY 416 YSRWESSNKALLEMAEKTVDKLEGLQVQIKQLEKLCALQTERDNLKRVQDLSAG 475
Db 244 YSRWESSNKALLEMAEKTVDKLEGLQVQIKQLEKLCALQTERDNLKRVQDLSAG 303
QY 476 GQGSITDGPERRPBPQGAQAPSSPRVTEAPCYCAPSTASGQTGPQEPSARA 530
Db 304 GQGSITDGPERRPBPQGAQAPSSPRVTEAPCYCAPSTASGQTGPQEPSARA 358

RESULT 13
AAB93594
ID AAB93594 standard; protein; 528 AA.
XX
AC AAB93594;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:13026.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000BP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00302523.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
PS Claim 8; SEQ ID NO 13026; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893

CC represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX
SQ Sequence 528 AA;
Query Match 47.5%; Score 1283.5; DB 4; Length 528;
Best Local Similarity 53.4%; Pred. No. 9.5e-72;
Matches 285; Conservative
QY 12 EGAQERSQAAPAVAEAGPSSQAPRK--PEGAQARTAQSGALEVDSEELSRQLEDILST 69
Db 7 EAARGGGGGAEEATEARGGRRSPRQKFRIGTWEEAGICGLGVKADMLCKNSQSDILQ 66
QY 70 YCVDNNGGPGEDGAOGEPAPEDAEKSRVTVARNGEPPTPVVYGEKPSKGPNTTEI 129
Db 57 Q--GSCNGGTSNKHSLSEDEGSDFITENRLVS-----PAYCTQESREIPGG---EA 114
QY 130 ROSDEVGDRDRHRPQEKKAAGLCKEITLLMQTLNTLSTPEKLAALCKKYAEILLEHRN 189
Db 115 RTDPPDGQDQDSECNRNKEKT--LGKEVLLMQALNTLSTPEKLAALCKKYADLESRS 172
QY 190 SOKMKLLOKKSOLVOEKDHLGHEHSHKAVIARSKLSLCLRELQHNRSLEBEGVQARRE 249
Db 173 VQKMKILQKKAQIVKEVHLQSEHSHKAVIARSKLSLCLRELQHNRSLEBEGVQARRE 232
QY 250 EEEKKEVTSHFQVTLNDIQLOMEQHNRSNKLQENNELAERLKKLIEQYELREHIDK 309
Db 233 EEEKKEATAHFQITLDEIQALQEQHDIHNAKLQENIELGEKLLKLEQYALREHIDK 292
QY 310 VFHKDLOOQLVDKLOQAQEMLKEAERHOREKDFLLKEAVESORMCMLKQEQETHLQ 369
Db 293 VFKRKLQOQLVDKLOQTTLKEADEKHQREHREFLKEATEHSHKYEQMKQEQEVOLKQ 352
QY 370 QLALYTFEFQNTLSKSGSEVFTTFQEMKMKTKKIKKLEKETTMYRSWESSNKALLE 429
Db 353 QLSLYMDKFEFFQNTLSKSGSEVFTTFQEMKMKTKKIKKLEKETTMYRSWESSNKALLE 412
QY 430 MAEKTVDKLEGLQVQIKQLEKLCALQTERDNLKRVQ-----DL 472
Db 413 MAEKTVDKLEGLQVQIKQLEKLCALQTERDNLKRVQ-----DL 472
QY 473 SAGGQGSIT--DSGPE-----RRPEGQAQA--PSSPR--VTEAPCYCAPSTEA 516
Db 473 ATPVWQPTALDSDSHKELNTSSKRALGAHLEAPKRSQSAVQKPESTGSAPIAES 526

RESULT 14
ABO07222
ID ABO07222 standard; protein; 395 AA.
XX
AC ABO07222;
XX
DT 13-AUG-2003 (first entry)
XX
DE Human p53 modifying protein, SBQ ID 182.
XX
KW Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic; antipapillary; p53 pathway; breast cancer; colon cancer; kidney cancer; lung cancer; ovarian cancer; angiogenesis; cell cycle; apoptotic disorder; cell proliferation disorder.
XX
OS Homo sapiens.
XX
PN WO2002299122-A1.
XX
PD 12-DEC-2002.
XX
PF 03-JUN-2002; 2002MO-US017382.
XX
PR 05-JUN-2001; 2001US-0296076P.
PR 10-OCT-2001; 2001US-0328605P.
PR 15-FEB-2002; 2002US-0357253P.

[illegible]

DB 303 KLICRALQTERNELKRVKVEVLKQVSIKAAIKAAKRDLATFVWQPCYALDSHKLMTSSAK 360

QY 489 PEGGQAQA-PSSPR--VTEAPCYGAPSTEA 516

DB 363 ALGAHLEAPKQSAVQKPFSTGSAPAISS 393

RESULT 15

ABR43284

ABR43284 standard; protein; 684 AA.

XX ABR43284;

XX AC

XX DT 09-JUL-2003 (first entry)

XX DE Human neurotransmission-associated protein NTRAN-14 SEQ ID NO:14.

XX KW Human; neurotransmission-associated protein; NTRAN; cytosolic; anti-HIV; antiarteriosclerotic; anticoagulant; neurotropic; neuroprotective; stroke; cerebroprotective; anti-allergic; anti-inflammatory; thyromimetic; cancer; antidiabetic; gene therapy; cell proliferative disorder; atherosclerosis; neurological disorder; epilepsy; Huntington's disease; immune disorder; inflammatory disorder; AIDS; allergy; developmental disorder; diabetes; hypothyroidism; Cushing's syndrome; endocrine disorder; infection.

XX OS Homo sapiens.

XX WO2003025129-A2.

XX PN 27-MAR-2003.

XX PD

XX PF 12-SEP-2002; 2002WO-US029219.

XX PR 14-SEP-2001; 2001US-0322180P.

XX PR 28-SEP-2001; 2001US-0326096P.

XX PR 04-OCT-2001; 2001US-0327446P.

XX PR 26-OCT-2001; 2001US-0345837P.

XX PR 02-NOV-2001; 2001US-0343903P.

XX PR 27-NOV-2001; 2001US-0334020P.

XX PR 07-DEC-2001; 2001US-0340226P.

XX PR 04-JAN-2002; 2002US-0345008P.

XX PR 18-MAR-2002; 2002US-0365645P.

XX PR 10-MAY-2002; 2002US-0379887P.

XX (INCY-) INCYTE GENOMICS INC.

XX Honchell CD, Warren BA, Borowsky ML, Griffin JA, Li JX, Lee SY; Yue H, Forsythe IJ, Marquis JP, Gletsen KJ, Baughtn MR, Tran UK, Lehr-Mason PM, Tang YJ, Ramkumar J, Emerling BM, Lee EA, Elliott VS; Hafalia AJA, Duggan BM, Chawla NK, Kable AE, Chang H, Khare R; Becha SD, Jin P, Lee S;

XX WPI; 2003-363137/34.

XX N-PSDB; ACC68992.

XX New human neurotransmission associated proteins (NTRAN), useful for diagnosing, treating and preventing diseases or conditions associated with the aberrant NTRAN expression e.g. cancer, AIDS, diabetes, epilepsy, or infections.

XX Claim 1; Page 201-203; 240pp; English.

XX ACC68979 to ACC69003 encode the human neurotransmission-associated proteins given in ABR43271 to ABR43295, designated NTRAN-1 to NTRAN-25 (1) (I) have cytosolic, antiarteriosclerotic, anticoagulant, neurotropic, neuroprotective, cerebroprotective, anti-HIV, anti-allergic, anti-inflammatory, thyromimetic and antidiabetic activities, and can be used in gene therapy. The NTRAN polypeptides and polynucleotides are useful in diagnosing, treating and preventing diseases or conditions associated with the decreased expression or overexpression of NTRAN, such as cell proliferative (e.g. cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS, allergies), developmental (e.g. hypothyroidism, Cushing's syndrome) or

CC endocrine (e.g. diabetes) disorders, or infections. They are also useful
CC in assessing the effects of exogenous compounds on the expression of
CC nucleic acid and amino acid sequences of NTRAN. The NTRANs or their
CC fragments are useful in screening compounds for effectiveness as agonist
CC or antagonist of the polypeptides, or in altering the expression of the
CC target polynucleotide and compounds that specifically bind to or modulate
CC the activity of the polypeptide
XX
XX Sequence 684 AA;

Query Match 46.7%; Score 1260.5; DB 6; Length 684;
Best Local Similarity 46.3%; Pred. No. 3.5e-70;
Matches 272; Conservative 91; Mismatches 135; Indels 89; Gaps 9;
QY 4 PQGPAGP--EQAQRPSCAAPAVEAGPGSCQAAPKPEGAQAQAQAGALRDVSEELSR 61
DB 18 PGSSSLPHNGLEKEDGQ-----DSPTVPQPEK-----BASVHPDISEELNR 61
QY 62 QLEDILTYCVNNGGPGEDGAGQGPAPPEPAEPAEAKSRRTYVARNGEPPTVPVYGEKPSK 121
DB 62 QLEDIINTY--GSAASTAGKGSARASEQENAESPDN---EDGCEETTEAGREPVAS 116
QY 122 GPNTPEIRIQSDVGRDHRHRRPQKKAGLKGKAITLLMTNTLSTPEKLAALCKKYA 181
DB 117 GPPTVTK-----BPVSKEQK--LEKILKGLGKEANLLMNKLTQPEKDFPKKYA 170
QY 182 ELLEHRSNOKMKLLQKQSVQVQKDLHRLGHSKAVLARSKLSLCSLRELOHNRSLKE 241
DB 171 ELLDEHRTQKLLKLLQKQVQIQEKDQLOQGEHSRAILARSKLSLCSLRELOHNRSLKE 230
QY 242 EGVORAREEERKKEVTSFQVTLNDILOQMEHNSKLRQENNELAERLKLIEQYE 301
DB 231 EALQAREEERKKEVTSFQVTLNDILOQMEHNSKLRQENNELAERLKLIEQYE 290
QY 302 LREEHIDKVFHKDLQQLVDAKLOQAQEMLKEAEERHOREKDFLLKEAVESQRMCELA 361
DB 291 LREEHIDKVFHKDLQQLVDAKLOQAQEMLKEAEERHOREKDFLLKEAVESQRMCELA 350
QY 362 QQETHLKQALALYTKPEEFQNTLSKSEVFTTTPKQEMEKMTKIKKLEKETTMYR 421
DB 351 EQETVLQALTLVSGRFEFQSTLTNSNEVPATFKQEMDKTKKMKKLEKDTATWKARPE 410
QY 422 SSKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSIT 481
DB 411 NCKALLDMIEEKAARAKEYECFVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSIT 470
QY 482 DSGPERRPEG-----DAEYNSVQAVYNLATAPMIIHPSTPHQSKETQPEIGSS 491
DB 471 QHNSDEEPSNSVSDQIEDAEYNSVQAVYNLATAPMIIHPSTPHQSKETQPEIGSS 530
QY 492 -----PGAQAPSP-----RVTEAPCVPGAPSTAESGQTGPQPTSA 528
DB 531 QESADAALKEPEQPPLIPSDSESPLPLTPQAEAGGSDAEPPSKA 577

RESULT 16
AAV17863
ID AAV17863 standard; protein; 676 AA.
XX AC AAV17863;
XX DT 16-AUG-1999 (first entry)
XX DE Neurite extending activity protein.
XX XX Neurite extending activity; anti-dementia; memory; brain function;
XX XX dementia.
XX OS Mus sp.
XX XX JF11147897-A.
XX XX 02-JUN-1999.

XX 13-NOV-1997; 97JP-00331242.
XX 13-NOV-1997; 97JP-00331242.
XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
XX WPI; 1999-379889/32.
XX N-PSDB; AAX80156.
XX New protein with neurite extending activity - useful for treating
XX dementia.
XX Claim 1; Page 5-8; 9pp; Japanese.
XX The present sequence represents a protein (I) which has neurite extending
XX activity. (I) may be administered to patients to prevent dementia or to
XX improve memory and brain function. (I) is also used as a neurite-
XX extending agent
XX Sequence 676 AA;
Query Match 45.6%; Score 1233; DB 2; Length 676;
Best Local Similarity 53.9%; Pred. No. 1.8e-68;
Matches 255; Conservative 84; Mismatches 100; Indels 34; Gaps 7;
QY 2 SSPQEPAGPAGAPRPSQAAPAVEAGPGSCQAAPKPEGAQAQAQAGALRDVSEELSR 61
DB 15 TSPTQDNQG-----QSKAEPVPVSPQLSP--TNQTSAPPEMATC-----DISEELNR 59
QY 62 QLEDILTYCVNNGGPGEDGQ-----QCEPAEPEDAERSRTYVARNGEPPTVPVYGEK 117
DB 60 QLEDIITY-----GSAASLVEKEGTTAETDKPEKEDVGSMEDEACEDVNESEKD 110
QY 118 BPSKDPNTEIRIQSDVGRDHRHRRPQKKAGLKGKAITLLMTNTLSTPEKLAALC 177
DB 111 KPAPGDAS-----RAKPSASKEQK--LEKILKGLGKEATLLMQSLNKLTPPEKLDLLF 164
QY 178 KYAELLEHRNSOROMKLLQKQSVQVQKDLHRLGHSKAVLARSKLSLCSLRELOHNR 237
DB 165 KYAELLEHRNSOROMKLLQKQSVQVQKDLHRLGHSKAVLARSKLSLCSLRELOHNR 224
QY 238 SLKEEGVORAREEERKKEVTSFQVTLNDILOQMEHNSKLRQENNELAERLKLIE 297
DB 235 TLKEETIQAREEERKKEVTSFQVTLNDILOQMEHNSKLRQENNELAERLKLIE 284
QY 298 BOYELREHIDKVFHKDLQQLVDAKLOQAQEMLKEAEERHOREKDFLLKEAVESQRM 357
DB 285 BOYELREHIDKVFHKDLQQLVDAKLOQAQEMLKEAEERHOREKDFLLKEAVESQRM 344
QY 358 ELKQOQETHLKQALALYTKPEEFQNTLSKSEVFTTTPKQEMEKMTKIKKLEKETTMYR 417
DB 345 XYLKQOETHLKQALALYTKPEEFQNTLSKSEVFTTTPKQEMEKMTKIKKLEKETTMYR 404
QY 418 SRWESSNALLLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQ 470
DB 405 SRFENCNALLDMIEEKAARAKEYECFVKIQRLEKLCRALQTERNDLNKRVQ 457

RESULT 17
AAW37883
ID AAW37883 standard; protein; 386 AA.
XX AC AAW37883;
XX DT 28-AUG-1998 (first entry)
XX DE BRC1A1 modulator protein 091-132Q20.
XX XX BRC1A1 modulator protein; 091-132Q20; breast cancer antigen 1;
XX XX tumour suppressor protein; diagnosis; therapy; human.
XX OS Homo sapiens.


```
XX FH Key Location/Qualifiers
XX FT 124. .143
XX FT /note= "leucine zipper motif"
XX PN WC9810066-A1.
XX XX
XX XX 12-MAR-1998.
XX XX
XX XX 06-AUG-1997; 97WO-US013944.
XX XX
XX XX 04-SEP-1996; 96US-0025601P.
XX XX
XX XX (ONYX-) ONYX PHARM INC.
XX XX
XX XX Rubinfield B, Polakis P, Ligenfelter C, Vuong TT;
XX XX
XX XX WPI; 1998-193616/17.
XX XX
XX XX N-PSDB; AAV29064.
XX XX
XX XX Breast cancer antigen 1 modulator protein - useful for diagnosing
XX XX diseases involving unwanted cell growth, e.g. breast cancer, and for
XX XX producing therapeutics for treatment of such diseases.
XX XX
XX XX Example 1; Fig 3; 73pp; English.
XX XX
XX XX This polypeptide comprises a 46 kDa BRCA1 modulator protein that binds to
XX XX the tumor suppressor gene product BRCA1, and which is characterised by a
XX XX leucine zipper motif. Its amino acid sequence was deduced from the
XX XX nucleotide sequence of a cDNA clone (see AAV29064), designated 031-132Q20
XX XX (ATCC 98143), isolated from a HeLa cell cDNA library using a yeast two-
XX XX hybrid assay. 3 cDNA clones (see also AAV29062 and AAV29063) coding for
XX XX BRCA1 modulator proteins (see AAV37881-83) have been characterised.
XX XX Vectors and host cells comprising the isolated nucleic acid sequences are
XX XX claimed, as well as a process for producing BRCA1 modulator protein by
XX XX culturing these host cells. BRCA1 modulator proteins and nucleic acids
XX XX can be used to diagnose diseases involving unwanted cell growth, e.g.
XX XX breast cancer, and to identify compounds that alter BRCA1 interaction
XX XX with BRCA1 modulators for the treatment of such diseases
XX XX
XX XX Sequence 386 AA;
XX XX
XX XX Query Match 38.4%; Score 1038.5; DB 2; Length 386;
XX XX Best Local Similarity 56.0%; Pred. No. 1.2e-56;
XX XX Matches 225; Conservative 55; Mismatches 99; Indels 23; Gaps 6;
XX XX
XX QY 13 GAQRPSPQAPAVAEAGPGSSQAPRK--PEGAQARTAQSGALRDVSEELSRQLEDILSTY 70
XX DB 4 GAEE-----ATEAGRGRRSPRQKFEIGTWEEAGICGLGVKADMLCNSQSDNLIHQH 56
XX QY 71 CVDNNGGPGEDGAGQEPEDAEKSRITYVARGSEPEPTPVYVGEKPSKGDPTNTEIR 130
XX DB 57 --GSGCGGTSNKHSLEEDGSDFTENRLVS-----PAYCTQESREIPGG----EAR 104
XX QY 131 QSDVGDGRDHRPQEKKAGLGEITLLMOTLNTLSTPEKLAALCKKYAELEEHNS 190
XX DB 105 TDPDQGDSECNKKT--LGEVILLMQALNTLSTPEKLAALCKKYADLLESRSV 162
XX QY 191 QKMKLQKQSQVLQVQKHURGHGSHKAVLARSKLSLCRELQHNRSLSKEGVQVQAREE 250
XX DB 163 QKMKLQKQQAIVQKVKHLSHSHKAILARSKLSLCRELQHNRSLSKEGVQVQAREE 222
XX QY 251 EEKKEVTSHPQVTLNDIQLQMBQHNERNKLRQENMELAEKLLIQVYELREHEDKV 310
XX DB 223 EERIEATAHQITLNEIQAOEQBDIHNAKLRENELGEKLLIQVYALREHEDKV 282
XX QY 311 FAKHDLQQLVDALQQAQEMKLABERHQBKDFLLKAEVDSORMCELMKQVQLKQ 370
XX DB 283 FKHRELQQLVDALQQLQTTQLKEADKQREFFLLKAEVDSORMCELMKQVQLKQ 342
XX QY 371 LALYTEKEEFQNTLSKSEVFTFKQEMEQTKIKKLEKE 412
XX DB 343 LSLYNDKEFEFTTWAKSNELFTFRQEMEQTKIKKIKKCKK 384
```

```
RESULT 18
AAV30151
ID AAV30151 standard; protein; 386 AA.
```

```
XX AAV30151;
```

```
AC AAV30151;
```

```
DT 27-OCT-1999 (first entry)
```

```
DE Amino acid sequence of a BRCA1 modulator protein.
```

```
KW Modulator protein; BRCA1; tumour suppressor protein; breast cancer;
```

```
KW ovarian cancer; cell growth; cell proliferation.
```

```
OS Homo sapiens.
```

```
XX Key Location/Qualifiers
```

```
FT Region 125. .143
```

```
FT /note= "leucine zipper motif"
```

```
XX US5948643-A.
```

```
XX 07-SEP-1999.
```

```
XX 13-AUG-1997; 97US-00968751.
```

```
XX 13-AUG-1997; 97US-00968751.
```

```
XX (ONYX-) ONYX PHARM INC.
```

```
XX Rubinfield B, Ligenfelter C, Vuong TT, Polakis PG;
```

```
XX WPI; 1999-517952/43.
```

```
XX N-PSDB; AAV86756.
```

```
XX Modulator proteins that bind to and modulate the activity of the BRCA1
```

```
XX tumor suppressor gene product, useful for the treatment of ovarian and
```

```
XX breast cancer.
```

```
XX Example 1; Fig 3; 35pp; English.
```

```
XX The present sequence represents a modulator protein, that binds to and
```

```
XX modulate the activity of the BRCA1 gene product (BRCA1). The BRCA1
```

```
XX protein has been characterized as a tumour suppressor protein.
```

```
XX Alterations in the amino acid sequence of BRCA1 causes breast and ovarian
```

```
XX cancers by removing the controls on cell growth and proliferation.
```

```
XX Research has shown that different regions on the BRCA1 molecule have
```

```
XX different effects on cell growth and tumour suppression (e.g. full length
```

```
XX truncated BRCA1 has no effect on breast cancer cell growth but will
```

```
XX inhibit ovarian cancer cell growth). It has been suggested that different
```

```
XX host cell factors (e.g. proteins) interact with different regions of the
```

```
XX BRCA1 to control its function. The identification of these proteins (e.g.
```

```
XX BRCA1MP) will facilitate the development of novel diagnostic methods and
```

```
XX new therapeutics for identifying and treating cancers caused by changes
```

```
XX in the expression or activity of BRCA1
```

```
XX Sequence 386 AA;
```

```
Query Match 38.4%; Score 1038.5; DB 2; Length 386;
```

```
Best Local Similarity 56.0%; Pred. No. 1.2e-56; Indels 23; Gaps 6;
```

```
Matches 225; Conservative 55; Mismatches 99; Indels 23; Gaps 6;
```

```
QY 13 GAQRPSPQAPAVAEAGPGSSQAPRK--PEGAQARTAQSGALRDVSEELSRQLEDILSTY 70
```

```
DB 4 GAEE-----ATEAGRGRRSPRQKFEIGTWEEAGICGLGVKADMLCNSQSDNLIHQH 56
```

```
QY 71 CVDNNGGPGEDGAGQEPEDAEKSRITYVARGSEPEPTPVYVGEKPSKGDPTNTEIR 130
```

```
DB 57 --GSGCGGTSNKHSLEEDGSDFTENRLVS-----PAYCTQESREIPGG----EAR 104
```

```
QY 131 QSDVGDGRDHRPQEKKAGLGEITLLMOTLNTLSTPEKLAALCKKYAELEEHNS 190
```


Db 105 TDPPDQDSECNNEKT--LGKEVLLMQALNTLSTPEKLAALCKKYADLLSESRV 162
QY 191 QKMKLLQKKQSQLVQEKDHLRGEHSAVLARSKLESCLREIQNRNRSUKRGVQARAE 250
Db 163 QKMKILQKKQQAIVKEVHLOSEHSAKILARSKLESCLREIQNRNKTLENNMQARAE 222
QY 251 ESKRKEVTSHPVLTNDIOLQMEHNRNSKLROENMELAEKLIQYELREHIDKV 310
Db 223 ERRIETATAHFQITLNEIQALQEHDIHNAKLQENIELGKLLIEQYALREHIDKV 282
QY 311 FKHKDLOQLVDKLAQQAQEMKAEERHOREKDFLKEAVESQRMCELMKQOETHLQK 370
Db 283 FKHEKLOQLVDKLAQQTQLIKEADEKHOREFLKEATESRHKYEQMKQEQVQLKQ 342
QY 371 LALYTEKFEFQNTLSKSEVFTTFKQEMERTKIKKLEKE 412
Db 343 LSLYMDKFEFQNTMAKSNELFTTFQEMERTKIKKXXXX 384

RESULT 19
AAB63259
ID AAB63259 standard; protein; 204 AA.
XX
AC AAB63259;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human breast cancer associated antigen protein sequence SEQ ID NO:621.
XX
DE Human breast cancer; gastric cancer; prostate cancer; diagnosis;
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
XX
KW cancer associated antigen; cytostatic; cancer vaccine.
XX
OS Homo sapiens.
XX
PN WO200073801-A2.
XX
PD 07-DEC-2000.
XX
XX 26-MAY-2000; 2000WO-US014749.
XX
PR 28-MAY-1999; 99US-0136526P.
PR 10-SEP-1999; 99US-0153454P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
XX Obata Y;
XX
XX WPI; 2001-025274/03.
XX
XX Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer.
XX
PS Example 1; Page 486-487; 799pp; English.
XX
XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC reduction of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer
XX
SQ Sequence 204 AA;

Query Match 30.2%; Score 815; DB 4; Length 204;
Best Local Similarity 76.0%; Pred. No. 4.5e-43;
Matches 155; Conservative 30; Mismatches 19; Indels 0; Gaps 0;

QY 163 LNTLSTPEKLAALCKKYAELEHNRNSQKMKLLQKKQSQLVQEKDHLRGEHSAVLAR 222
Db 1 LNTLSTPEKLAALCKKYADLLSESRVQKMKILQKKQQAIVKEVHLOSEHSAKILAR 60
QY 223 SKLESCLREIQNRNRSUKRGVQARAEERKEVTSHPVLTNDIOLQMEHNRNSKL 282
Db 61 SKLESCLREIQNRNKTLENNMQARAEERKEVTSHPVLTNDIOLQMEHNRNSKL 120
QY 283 ROENMELAEKLIQYELREHIDKVFKHDKLOQLVDKLAQQAQEMKAEERHORE 342
Db 121 ROENIELGKLLIEQYALREHIDKVFKHDKLOQLVDKLAQQAQEMKAEERHORE 180
QY 343 KDFLLKEAVESQRMCELMKQOETH 366
Db 181 REFLLKEATESRHKYEQMKQEQVH 204

RESULT 20
AAO08301
ID AAO08301 standard; protein; 475 AA.
XX
AC AAO08301;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 22193.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US004927.
XX
PR 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX N-PSDB; AAI88232.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
XX Claim 20; SEQ ID NO 22193; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 475 AA;

Query Match 30.0%; Score 810; DB 4; Length 475;
Best Local Similarity 47.7%; Pred. No. 2.5e-42;

Human; diagnostic; drug screening; forensics; gene mapping; biodiversity assessment; Parkinson's disease; Alzheimer's disease; neurodegenerative diseases; anaemia; platelet disorder; wound; burns; ulcers; osteoporosis; autoimmune disease; cancer; molecular weight marker; food supplement; antiparkinsonian; neurotropic; neuroprotective; antianemic; anticoagulant; thrombolytic; vulnerary; antitumor; osteopathic; immunosuppressive; antiinflammatory; cytostatic; gene therapy; chromosome 6.

OS Homo sapiens.

XX WO2003029271-A2.

XX 10-APR-2003.

XX 24-SEP-2002; 2002WO-US030474.

XX 24-SEP-2001; 2001US-0324631P.

XX (HYSE-) HYSEQ INC.

XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T; Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G; Haley-Vicente D, Drmanac RT;

XX WPI; 2003-371981/35.

XX N-PSDB; ADC30642.

XX New polynucleotide and polypeptide useful for diagnosing, preventing or treating conditions such as neurodegenerative diseases, anemias, platelet disorders, wounds, ulcers, osteoporosis, autoimmune diseases or cancer.

XX Claim 20; SEQ ID NO 1695; 1185pp; English.

XX The invention relates to 971 novel human cDNA sequences (ADC29919-ADC30889) and the polypeptides they encode (ADC30890-ADC31850). The invention also relates to nucleic acid sequences over 99% identical with the novel human cDNAs. The invention additionally encompasses expression vectors and host cells comprising a nucleic acid of the invention; the recombinant production of a polypeptide of the invention; an antibody against a polypeptide of the invention; a method of detecting polynucleotides or polypeptides of the invention; and methods of identifying a compound which binds to a polypeptide of the invention. The invention further discloses methods of preventing, treating or ameliorating a medical condition; kits comprising polynucleotide probes and/or monoclonal antibodies for carrying out the methods of the invention; methods for the identification of compounds that modulate the expression or activity of the polynucleotide and/or polypeptide; and 767 cDNA sequences corresponding to the cDNA sequences of the invention (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628-ADC33394). The nucleic acids and polypeptides of the invention are useful in diagnostics, drug screening, forensics, gene mapping, in the identification of mutations responsible for genetic disorders or other traits, for assessing biodiversity, and in producing many other types of data and products dependent on DNA and amino acid sequences. They are also used for treating diseases such as Parkinson's disease, Alzheimer's disease and other neurodegenerative diseases, anaemia, platelet disorders, wounds, ulcers, osteoporosis, autoimmune diseases or cancer. The nucleic acids may also be used as hybridisation probes or primers, and in the recombinant production of a protein. The polypeptides are also useful in generating antibodies, as molecular weight markers, and as food supplements. The present sequence represents a specifically claimed human polypeptide sequence of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 416 AA;

Query Match 24.4%; Score 659; DB 7; Length 416;
Best Local Similarity 44.2%; Pred. No. 5.3e-33;
Matches 136; Conservative 50; Mismatches 62; Indels 60; Gaps 2;

QY 281 KLRQNNMELAEERLKKLIEQVELREHIDKVPFKHDLQOOLVDAKLOQAQEKLEABERHQ 340
DB 2 KLCQENTELAEKLSIIDQVELREHIDKIPKHELOQKLVDKLEQAQEKLEABERHK 61
QY 341 REKDFLKAQVESORMCELMKQOETHLKOOLALYTEKFEFQNTLSKSEVFTTFKQEME 400
DB 62 REKYLLNQAAEWKQAKVLEQIVLQAQTLTSGRFEFQSTLTKSNEVFAITFKQEMD 121
QY 401 XMTKKIKKLEKETTMYRSRMSSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQT 460
DB 122 KTTKKMKKLEKDTATKARFENCKALLDMTEKALRAKEVECFVMKIGRLNLCRALQE 181
QY 461 ERNDLNKRVQDLSAGGQSLTDSQFERRPEG----- 491
DB 182 ERNELKKKIRDAEISEKDDQSQHSNDEPEFNSVDOEIDAEVNSVQTAVNLTAFMI 241
QY 492 -----PCAQAPSSP-----RVTEAPCYPGAPSTASGQT 520
DB 242 IHPPESTPHQSKEQTEIGSSQBSADAALKEPEQPLIPSRDSESPLPLTPQAEAGGS 301
QY 521 GPQEPRTSA 528
DB 302 DAEPPSKA 309
RESULT 23
ABG02486
ID ABG02486 standard; protein; 410 AA.
XX AC ABG02486;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #2477.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS66673.
XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.
XX Claim 20; SEQ ID NO 32845; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABB00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 410 AA;

Query Match 23.8%; Score 642.5; DB 4; Length 410;
 Best Local Similarity 40.9%; Pred. No. 5.6e-32;
 Matches 153; Conservative 55; Mismatches 85; Indels 81; Gaps 11;

QY 167 STPEKLAALC-KKYABLL-EEHNSQKQKMLQKQS--QLVQKHLRGEHSAV 219
 DB 85 STMEE--AGLCGLREKADMLNCSSEH-----DILQHQDSCSATSNNKHLLEDEGRDF 135

QY 220 LARSKLESLELQNRNLSKEGVORAREEKEKVTSHFQVTLNDIQLQMEQHNRN 279

DB 136 IYK-----NRSW-BEHMQCKKEEVVKEVTAHFQITLTQALQEQHEIHN 181

QY 280 SKLRQENELARLKLQYBLREHIDKVPKHKDLOOQLVDKQLQQAQEMKLEAERH 339

DB 182 AKLQENEMEGEKLKLTQVALREBQINKAFKHLRQLVLDARLQTAQLINEADERH 241

QY 340 QREKDFLLKEAVESORMCBLKQOETHLQOALVTERKPEFQNTLSKSSVEFTTQDEM 399

DB 242 QREREFLLKEAVESORMCBLKQOETHLQOALVTERKPEFQNTLSKSSVEFTTQDEM 301

QY 400 EMTKKIKKLEKETTGYRSRWSSNNKALLEMAHEKTVTRDKLEGLQVKIQLEKLCRALQ 459

DB 302 EK-----XTIRDKNYKVPQIKLERLEKLYKALQ 329

QY 460 TERNDLNKRV-----QDLSAGGQGLTDSGPER-----RPEGPAQAPSSR 501

DB 330 IERNELSEKGLTKGVSVKQVADVLAVPVTHSCADLSSNMKNTSSKRAFGVLEADPK 369

QY 502 -VTEAPCYGAPST 514

DB 390 GMEVCKYKALST 403

RESULT 24

ABB61894

ID ABB61894 standard; protein; 515 AA.

XX AC ABB61894;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 12474.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW Pharmaceutical.

XX OS Drosophila melanogaster.

XX FN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX FR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE) PE CORP NY.

XX PI

XX DR Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL05997.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX PT genes from Drosophila and for elucidating cell signalling and cell-cell

XX PT interactions.

XX PS Disclosure; SEQ ID NO 12474; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent

XX CC capable of detecting 1000 or more genes from Drosophila. The invention is

XX CC useful in developmental biology and in elucidating cell signalling and

XX CC cell-cell interactions in higher eukaryotes for the development of

XX CC insecticides, therapeutics and pharmaceutical drugs. The invention

XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

XX CC ABB72072). The sequence data for this patent did not form part of the

XX CC printed specification, but was obtained in electronic format directly

XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 515 AA;

Query Match 22.1%; Score 597.5; DB 4; Length 515;

Best Local Similarity 35.8%; Pred. No. 4.6e-29;

Matches 126; Conservative 98; Mismatches 105; Indels 23; Gaps 3;

QY 127 BEIROSDEVDGDRHRRPQKAKGLGKEITLMTQTLNLTSPPEKLAALCKYAEILEE 186

DB 2 EKLKAKKAVAREKORDKLE-----LVMSLDECPSAEKVKL-----LLQR 45

QY 187 HNSOK-----QMKLQKQSQVLQVQKHLRGEHSAVLAARSLKSLRELQHNHSL 239

DB 46 HYDSEKNVSRLLFAELRVLQRMESQOREKEQVQRLNKLNRDLQGVCEQOQRIKSV 105

QY 240 KEEGVORAREEKEKVTSHFQVTLNDIQLQMEQHNRNKLQENNELAERLKLIEQ 299

DB 106 KNESSLQIKVBERREKESQTKFOSSLNDVQKSLAGNENIKLDYNIEMTKLKLAEQ 165

QY 300 YELREHIDKVPKHKDLOOQLVDKQLQQAQEMKLEAERHOREKDFLKEAVESORMCEL 359

DB 166 YQTRQHLEKLENEQVLEAQLHQAQKQRCQVEAAAMEKILSKENQIGLEKLMQAQRAIKD 225

QY 360 MKQOETHLQOALVTERKPEFQNTLSKSSVEFTTFOEMKMTKIKKLEKETTMYRSR 419

DB 226 LTDRHQLEKELNIYTKYDDFQQLQKSNVFGSYKVELKMSKGTCKIEKALGNRQK 285

QY 420 WESSNKALLEMAHEKTVTRDKLEGLQVKIQLEKLCRALQTERNDLNKRVQD 471

DB 286 YEKANAMVIDLATERSLQTHSERLQKQIQOQLKLLRALQLERTTLHKCLRD 337

RESULT 25

ABG02487

ID ABG02487 standard; protein; 841 AA.

XX AC ABG02487;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #2478.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX FN WO200175067-A2.

XX PD 11-OCT-2001.

	30-MAR-2001; 2001WO-US008631.	
PPF		
XX		
PER	31-MAR-2000; 2000US-00540217.	
XX	23-AUG-2000; 2000US-00649167.	
PPR	(HYSE-) HYSEQ INC.	
PAP		
FFI	Dmanac RT, Liu C, Tang YT;	
XX	WPI; 2001-639362/73.	
DER	N-PDSB; AAS66674.	
XX		
PT	New isolated polynucleotide and encoded polypeptides, useful in	
PT	diagnostics, forensics, gene mapping, identification of mutations	
PT	responsible for genetic disorders or other traits and to assess	
PT	biodiversity.	
XX		
PS	Claim 20; SEQ ID NO 32846; 103pp; English.	
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)	
CC	sequences. (I) is useful as hybridisation probes, polymerase chain	
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,	
CC	and in recombinant production of (II). The polynucleotides are also used	
CC	in diagnostics as expressed sequence tags for identifying expressed	
CC	genes. (I) is useful in gene therapy techniques to restore normal	
CC	activity of (II) or to treat disease states involving (II). (II) is	
CC	useful for generating antibodies against it, detecting or quantitating a	
CC	polypeptide in tissue, as molecular weight markers and as a food	
CC	supplement. (II) and its binding partners are useful in medical imaging	
CC	of sites expressing (II). (I) and (II) are useful for treating disorders	
CC	involving aberrant protein expression or biological activity. The	
CC	polypeptide and polynucleotide sequences have applications in	
CC	diagnostics, forensics, gene mapping, identification of mutations	
CC	responsible for genetic disorders or other traits to assess biodiversity	
CC	and to produce other types of data and products dependent on DNA and	
CC	amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic	
CC	amino acid sequences of the invention. Note: The sequence data for this	
CC	patent did not appear in the printed specification, but was obtained in	
CC	electronic format directly from WIPO at	
CC	ftp.wipo.int/pub/published_pct_sequences	
XX		
SQ	Sequence 841 AA;	
	Query Match 20.4%; Score 550; DB 4; Length 841;	
	Best Local Similarity 50.0%; Pred. No. 7.4e-26;	
	Matches 121; Conservative 40; Mismatches 51; Indels 30; Gaps 7	
Oy	167 STPEKKAALC--KKYAEILL---EEHNSOKMKLLQKQS--QLVQEKDHLRGESXAV 219	
Db	53 STMSE--AGLCGLREKADMLNCSSEH-----DLIQHQDNCSATSNKHLEDEGRDF 103	
Oy	220 LARGSLIESLRELQRNHRSLKEGVQVARREEBKRXEVTSHFQVTLDITQLQMEQHNRN 279	
Db	104 ITK-----NRSW-EHHMQQKKEEVEVLKEVTAHTQIILTETQAQLQEHIHN 149	
Oy	280 SKLQENKELAERLKULIEQVELREEHKDKVFVKHQLQQOLVDAKLQQOAEMLKAEERH 339	
Db	150 AKLAQENMEMGEKLKCLTDQVALREEQINKAFPHKELRQOLVDARLQQQTAAQLKEADERH 209	
Oy	340 QREKDPLLKAEVSORMCELMKQOETHLKQCALYTEKFEFPONTLSKSSEVFTTFKOBM 399	
Db	210 QREREPFLKEATESRHKYEEKQBEAQLKEQLFLYMDKGFEEPOTTWAKTNELTFAFKSET. 269	
Oy	400 EK 401	
Db	270 EK 271	
RESULT 26		
ABG02488		
ID	ABG02488 standard; protein; 102 AA.	
XX		
AC	ABG02488:	

Db 565 KVAVSPTPPVSPVSTPPVAPPEQSEALAKMBAVAQVLE-----QDQHLLSKOE 619
QY 203 QLVQEKDHLRGEHSKAVLA--RSKLSLCLREIQHNRSLKEGVQARBEK-----RK 255
Db 620 KMQLREKLCQEBEIBILHQKEQSLSLRLQKATEEBA-RMRBEESQRLSWLRA 678
QY 256 EVTSHFQVTLNDIQLQVE-----QHNRSLKQENMELAEKLIQOYELR 303
Db 679 QVQSSQADQIRACQASLQKLRLESLSQKAPASLEQKRMQLKKEIEASEKS 738
QY 304 BEHIDKVFHKLIQOQ---QLVDAKLQQAQEMLE-----ABERHOREKDFLLKE 349
Db 739 EQAALNAAKKALQQLREQLGEGRKEAVATLEKHSABLERLCSLEAKHREVSSLOKK 798
QY 350 AVESQRMCLMKQ---OETHLQQLALYTEKPEEFQNTL--SKSEVETTFKQEMKVT 403
Db 799 IOAQOQKEAQKQKCLGQVHRVHQSYHVAGYEHLSLLEKREKQEVGEHRRIDRM- 857
QY 404 KTKIKLEKXTMYRSWESSNKALLBMAEKTVDKXELGLOVKIQRLKLC-RALQTER 462
Db 858 ----KEHQQVMKAREQVE-----ABERKQASLLGHLTGELERLQRAHRELETVR 906
QY 463 NDLNKRVDL 472
Db 907 QEQHKELEDL 916

RESULT 30
ADC38517
ID ADC38517 standard; protein; 860 AA.
XX
AC ADC38517;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human AMLP1 myosin-tail motif consensus protein SEQ ID NO:866.
XX
KW human; angiominotin-like protein 1; AMLP1; cytostatic; gene therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO2003037931-A2.
XX
PD 08-MAY-2003.
XX
PF 01-NOV-2002; 2002WO-US035129.
XX
PR 01-NOV-2001; 2001US-0334773P.
XX
PA (AMSH) AMERSHAM BIOSCIENCES SV CORP.
XX
PI Shannon M, Phan T;
XX
PWPI; 2003-430501/40.
XX
PT New isolated nucleic acid molecule encoding a human angiominotin-like
PT protein, useful for treating or preventing a disorder associated with
PT decreased or increased expression or activity of AMLP1.
XX
PS Example 2; SEQ ID NO 866; 172pp; English.
XX
CC The present invention describes the human angiominotin-like protein 1
CC (AMLPI). human AMLPI has cytostatic activity and can be used in gene
CC therapy. The AMLPI protein, nucleic acid molecules, antibodies, and
CC compositions of the present invention can be used for treating or
CC preventing a disorder associated with decreased or increased expression
CC or activity of AMLPI. The present sequence represents a consensus myosin-
CC tail motif amino acid sequence, which is used in an example from the
XX present invention.
XX
SQ Sequence 860 AA;

Query Match 9.5%; Score 257.5; DB 7; Length 860;
Best Local Similarity 22.7%; Pred. No. 1.2e-07;
Matches 143; Conservative 106; Mismatches 189; Indels 193; Gaps 25;
QY 12 EGAERPSQAAPAVEAEGFGSSQAPRKEPGAQRTAQSGALRDVSELSRQLEDILSYTC 71
Db 77 EELSERLEAGGATAAQ-----IELNKKRAELAKLRKDLERANLQHEEALAT-L 125
QY 72 VDNQGGGEGDGAQGEPAEPEDAESRYYVARNGEPETPVVYGEKPSKGDPTNEB-IR 130
Db 126 RKQHODAINELSEIQLOKQAK-----LLMOTLNTLSTPEEKJAA- 166
QY 131 QSDEVDGRDHRPPOEKKAKGLGKEIT-----LLMOTLNTLSTPEEKJAA- 175
Db 167 QLDSI---TKAKNAEKAKQLESQLSQVLEKLBQSLNDITSQKRLQSENSDLTRQ 223
QY 176 -----LCKYAEI---LEHRNS-----QKMKLQKQSQVQKQHLRAGHSA 218
Db 224 LEBAQVSNLSKLQSQLESQLEAKGSLSEESERANLQALQLEHDLQSLRQLEEE 283
QY 219 VLAKSLSLQELQHRNRSIKE-----EGVQARBEERKEKVT----- 258
Db 284 SEAKAEI---RLSKANAEIQWRSKFESESGALRAEELBLKKLNQKISELEBAEAA 340
QY 259 -----SHFOVTINDIQLQVQHNERNKIL--ROENMB--LABRIKKLI-E----- 298
Db 341 NAKCDSLEKTKSLQSELEDLQIELERANAAASBLKKQKQKFDKILAEWKKQVDELQAE 400
QY 299 ---QVELR-----BEHIDKV-----FKHKLQOQVDAKLQQAQ-----EMLK 333
Db 401 DTAQREARNLSTELFKNLEELKQVLEALRENNKLNQDSIHDLTDQLGEGGRNVHELE 460
QY 334 BAEERHOREKDFLLKEAVESQRMCL-----MKQETHLQKQALALYTEKPEEPQ 382
Db 461 KARRLEAKDELOAALBEAALAEBSKVLRQAQVELSQIRSEIERLIA--EKBEFE 517
QY 383 NT-----LSKSSSEVETTFKQEMKMTKK-----IK 407
Db 518 MTRKNHQRAIESLOATLEATGKAEASRLKKLEGDINELEIALDHANKANAEAKQNVK 577
QY 408 KLEKETTMYRSWRSSNKA-----LLEMAEEK-TVRDKLEGLQVKTORLEKLCRAQOT 460
Db 578 KYQQQKVELQTVBEEQARADAREQLAVAEERRATALEABLEELSALEQAERARKQAE 637
QY 461 ERNDLNKQVQLSAGCGQSLTDSGPEREPG 491
Db 638 ELAEASERVNLTQNSSLIAQ---KKKLEG 665

RESULT 31
AAV07031
ID AAV07031 standard; protein; 1752 AA.
XX
AC AAV07031;
XX
DT 02-JUL-1999 (first entry)
XX
DE Breast cancer associated antigen precursor sequence.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US014679.
XX
PR 17-JUL-1997; 97US-00896164.
PR 10-OCT-1997; 97US-0061599P.

DB 765 REKSRKNSLSAEIERLQAEIKRIEBCRKLESTRETQSLTERSRVQBIDKL-- 823
 QY 475 GGQGSILTDGPPRRPPEGCAQAPSSPRVTEAPC 507
 DB 824 -----RQRPYG-----SHRETQTEC 838

RESULT 32
 ABU07402
 ID ABU07402 standard; protein; 2871 AA.
 AC ABU07402;
 XX 28-JAN-2003 (first entry)
 DE Protein differentially regulated in prostate cancer #3.
 KW Prostate cancer; gene expression; differential regulation;
 KW molecular marker; drug target; cancer detection; cancer diagnosis;
 KW cancer staging; cancer grading; cancer assessing; cancer monitoring.
 OS Homo sapiens.
 PN W0200281638-A2.
 PD 17-OCT-2002.
 PF 08-APR-2002; 2002WO-US010824.
 PR 06-APR-2001; 2001US-0281731P.
 PR 06-APR-2001; 2001US-0281732P.
 XX (ORIG-) ORIGENE TECHNOLOGIES INC.
 XX Sun Z, Jay G;
 DR WPI1; 2003-058520/05.
 PT Novel genes which are differentially regulated in prostate cancer, useful
 PT for diagnosing prostate cancer in prostate tissue sample and assessing
 PT therapeutic or preventive intervention in prostate cancer patients.
 XX Claim 1; Page 209-219; 416pp; English.

The invention describes genes (I) which are differentially regulated in prostate cancer. (I) is useful for diagnosing a prostate cancer in a sample comprising prostate tissue, which involves determining the number of target genes which are differentially-regulated in the sample, where the number is indicative of the probability that the sample comprises prostate cancer. (I) is useful for assessing a therapeutic or preventive intervention in a subject having a prostate cancer, which involves determining the expression levels in a sample comprising prostate tissue of target genes which are differentially-regulated in prostate cancer. Preferably, the expression levels of at least 10 genes are determined. (I) is also useful for identifying agents that modulate a biological activity of a polypeptide differentially-regulated in prostate cancer cells, which involves contacting a polypeptide differentially-regulated in prostate cancer cells with a test agent under conditions effective for the test agent to modulate a biological activity of the polypeptide, and determining whether the test agent modulates the biological activity. (I) is useful as molecular markers, as drug targets, and for detecting, diagnosing, staging, grading, assessing, monitoring, prognosticating, preventing or treating, determining predisposition to diseases and conditions especially relating to prostate cancer. (I) and its expression products are used in the diagnostic test to assay for presence of cancer e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in blood etc. (I) is useful for assessing cancer e.g., to determine the type of cancer, its stage of development, the nature of genetic defect, etc. The polypeptide encoded by (I) can be used as target for therapy or drug discovery. (I) can also be used for expressing the polypeptide and thus for searching specific binding partners of the polypeptide. (I) is useful in therapeutic applications to treat prostate cancer. The identification of specific genes, and groups of genes, expressed in pathways

PR 10-OCT-1997; 97US-0061765P.
 PR 10-OCT-1997; 97US-00948705.
 PR 11-OCT-1997; 97GB-00021697.
 PR 22-JUN-1998; 98US-00102222.
 XX (LUDW-) LUDWIG INST CANCER RES.
 XX Old LJ, Scanlan MJ, Stockert E, Gure A, Chen Y, Gout I;
 PI O'hare M, Obata Y, Pfeundschnuh M, Tureci O, Sahin U;
 XX WPI1; 1999-132448/11.
 XX New isolated cancer associated nucleic acids and polypeptides - isolated
 PT using sera from cancer patients, used to develop products for the
 PT diagnosis, monitoring or treatment of cancers.
 XX Disclosure; Page 409-413; 787pp; English.
 XX The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer
 XX Sequence 1752 AA;
 SQ

Query Match 9.5%; Score 256; DB 2; Length 1752;
 Best Local Similarity 23.2%; Pred. No. 3.5e-07;
 Matches 133; Conservative 107; Mismatches 195; Indels 138; Gaps 25;

QY 26 EAGPGSSQAPRP-----EGNARTAGS-----GALRDVSELSRQEDIL 67
 DB 313 QKQATSEVSQKQQLLEVRQVQMTBESVRYKQSLDAAKTIDKNKBIER-LKLI 371
 QY 68 STYCVNNNGGPDGAGPAPDAEAKSRT-YVARNGEPPEPTPVV---YGEKPSKG 122
 DB 372 DKEINR-----KLEDENARLQVQYDLQKANSATETINKLKVQSELTRL 419
 QY 123 DPNTETIRQDEVDHRPPOEKYKAGLGKEITLLMOTLNTSTPEEKLAALCKYAE 182
 DB 420 RIDYERSQERTVQDDITRFQNSLRELQKQ--KVEEELNRLKRTASDSCRKKEES 477
 QY 183 LLEHRNSQOMKL-----LQKQSQ--LVQSKD-----HLR-GEHSKAVLAR 222
 DB 478 ELEGRRSLKEQAIKNTLQQLQEQASIVKRSDEDLQQRDVLGHLRQKQTEBLR 537
 QY 223 --SKLSLCLRELQHNHSLKE-----EGVQARBEKEKKEVTSHFQVTLNDIQLQEQH 275
 DB 538 LSSEVEALRQLLQEQESVQKALNEHFQKATEDKSR-----SLNESKIBIERL 587
 QY 276 NERNKLRQENMELARLKL-TEOVELR-----BEHIDK-----VPKHKDLQ 318
 DB 588 QSLTENTKHELMEEELRLREYDLRGREASDKNATIELRSQIQISNNITL 647
 QY 319 QLVDAKLQQAQEMKEAEERHQEKDFLLKEAVESQRMCC-ELMKQETHLQQLALYTEK 377
 DB 648 QGLINDLQRENLRLQRIEFKQALASNRQIESKNQCTQVQVERESLLVKKVLEQDK 707
 QY 378 -----FEFQNTLKSSEVETTKQEMKTKKKK-LEKETTMYRSWSSNKALLEMA 431
 DB 708 ARQLRDELDENAKSTLEATRVKQRLCECKQIQNDLNQMTQVSKKEAIBRK--IESE 765
 QY 432 EEKTVRK-----ELEGLQVKIORLEKLCR-----ALQTERNDLNKRVODLSA 474

CC physiologically relevant to prostate cancer permits the definition of
CC functional and disease pathways and the delineation of targets in these
CC pathways which are useful in diagnostic, therapeutic, and clinical
CC applications. This is the amino acid sequence of a protein differentially
CC regulated in prostate cancer
XX
XX
SQ Sequence 2871 AA;

Query Match 9.5%; Score 256; DB 6; Length 2871;
Best Local Similarity 23.2%; Pred. No. 6.2e-07;
Matches 133; Conservative 107; Mismatches 195; Indels 138; Gaps 25;
QY 26 BAEGPGSQAPRKP-----EGQAARTAQ-----GALRDVSELSQLDIL 67
DB 1432 QKATGSEVSQKQOQLEVLQVQMTREESVRYKQSLDDAAKTIQDKNEIER-LKQLI 1490
QY 68 STYVDNNQGGPGEDGAGPAPEDAERT-YVARGPEPTPVV-----YGEKPSKG 122
DB 1491 DKETNDR-----KCLEENARLQVQVLDQKANSATETINKLVQEQELTRL 1538
QY 123 DPNTBEIRQSDVGDHRRPQEKKAAGLGEITLLMTLSTPEEKLAALCKKYAE 182
DB 1539 RIDYERSQERTVQDDITRFQNSLQELQKQ--KVEEELNRLKRTASEDSCKRKL 1596
QY 183 LLEHRNSQKQMKL-----LQKQSQ--LVQEKD-----HLR-GEHSAVLAR 222
DB 1597 ELEGRRSLKEQAIIKITNLTOLEQASIVKRSSEDDLQQRDVLGDHLEKQTOEELR 1656
QY 223 --SKLESCLRELQHNNSLKE-----EGVQARBEERKEKVTSHFQVTLNDILOMQE 275
DB 1657 LSSEVALRRQLQEQESVKQAHLEHNFQKAIEDKSR-----SLNESKIEIRL 1706
QY 276 NERNKSLRQENNELAERLKKL-IEQVELR-----BEHIDK-----VFKHDLQ 318
DB 1707 QSLTENTLKEHLMLEELNRLLEYDRLRRGRSEADSKNATILELSQLQISNNRTLEL 1766
QY 319 QLVDAKIQQAQEMLKEAERHOREKDFLLKEAVESQRC-ELMKQOETHLQQLALYTEK 377
DB 1767 QGLINDLQRENLQRENLQRENLQRENLQRENLQRENLQRENLQRENLQRENLQ 1826
QY 378 -----FEFPQNTLSKSSEVFTTFKQEMKMTKIKK-LEKETTYRWSWSSNKALLEMA 431
DB 1827 ARQLRLEDELNRAKSTLEAEITRVKQRLCEKQIQNDLQNMKTQYSRKEBAIRK--IESE 1884
QY 432 EKTVRDK-----ELEGVQKIQRLKLCR-----ALQTERNDLNKRVQDLSA 474
DB 1885 REKSERKNSLRSEIERLQAEIKRIEERCCKRLKEDSTRETOSQLETYSRYQREIDKL-- 1942
QY 475 GGQSLTDSGPERPEGPGCAQAPSPVTEAPC 507
DB 1943 -----RQRPYG-----SHRETQTEC 1957

RESULT 33
ADC35075
ID ADC35075 standard; protein; 2871 AA.
XX
XX AC ADC35075;
XX
XX DT 18-DEC-2003 (first entry)
XX
XX DE Human breast cancer antigen seq id 41.
XX
XX KW breast cancer; breast cancer diagnosis; breast cancer antigen.
XX
XX OS Homo sapiens.
XX
XX PN US2003108888-A1.
XX
XX PD 12-JUN-2003.
XX
XX PF 15-MAY-2002; 2002US-00146473.
XX

PR 15-MAY-2001; 2001US-0291150P.
XX
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX
XX PI Scanlan MJ, Gout I, Stockert B, Old LJ, Gure A, Chen Y;
XX
XX DR WPI; 2003-829397/77.
XX
XX DR N-PSDB; ADC35117.
XX
XX PT Diagnosing breast cancer in subject by obtaining biological sample from
XX
XX PT subject, contacting sample with breast cancer-associated polypeptides,
XX
XX PT determining specific binding between polypeptides and agents in sample.
XX
XX PS Example 2; SEQ ID NO 41; 173pp; English.
XX
XX CC The invention describes a method of diagnosing breast cancer in subject
XX
XX CC comprising contacting biological sample from subject with at least two
XX
XX CC different breast cancer-associated polypeptides (I) encoded by nucleic
XX
XX CC acid molecules (II) comprising sequence chosen from 42 fully defined
XX
XX CC sequences as given in specification, determining specific binding between
XX
XX CC (I) and agents in sample, where presence of the binding is diagnostic for
XX
XX CC breast cancer. The method is useful for diagnosing breast cancer in a
XX
XX CC subject. The sample is blood, lymph node fluid or breast discharge fluid.
XX
XX CC This is the amino acid sequence of a breast cancer antigen.
XX
XX SQ Sequence 2871 AA;

Query Match 9.5%; Score 256; DB 7; Length 2871;
Best Local Similarity 23.2%; Pred. No. 6.2e-07;
Matches 133; Conservative 107; Mismatches 195; Indels 138; Gaps 25;
QY 26 BAEGPGSQAPRKP-----EGQAARTAQ-----GALRDVSELSQLDIL 67
DB 1432 QKATGSEVSQKQOQLEVLQVQMTREESVRYKQSLDDAAKTIQDKNEIER-LKQLI 1490
QY 68 STYVDNNQGGPGEDGAGPAPEDAERT-YVARGPEPTPVV-----YGEKPSKG 122
DB 1491 DKETNDR-----KCLEENARLQVQVLDQKANSATETINKLVQEQELTRL 1538
QY 123 DPNTBEIRQSDVGDHRRPQEKKAAGLGEITLLMTLSTPEEKLAALCKKYAE 182
DB 1539 RIDYERSQERTVQDDITRFQNSLQELQKQ--KVEEELNRLKRTASEDSCKRKL 1596
QY 183 LLEHRNSQKQMKL-----LQKQSQ--LVQEKD-----HLR-GEHSAVLAR 222
DB 1597 ELEGRRSLKEQAIIKITNLTOLEQASIVKRSSEDDLQQRDVLGDHLEKQTOEELR 1656
QY 223 --SKLESCLRELQHNNSLKE-----EGVQARBEERKEKVTSHFQVTLNDILOMQE 275
DB 1657 LSSEVALRRQLQEQESVKQAHLEHNFQKAIEDKSR-----SLNESKIEIRL 1706
QY 276 NERNKSLRQENNELAERLKKL-IEQVELR-----BEHIDK-----VFKHDLQ 318
DB 1707 QSLTENTLKEHLMLEELNRLLEYDRLRRGRSEADSKNATILELSQLQISNNRTLEL 1766
QY 319 QLVDAKIQQAQEMLKEAERHOREKDFLLKEAVESQRC-ELMKQOETHLQQLALYTEK 377
DB 1767 QGLINDLQRENLQRENLQRENLQRENLQRENLQRENLQRENLQRENLQRENLQ 1826
QY 378 -----FEFPQNTLSKSSEVFTTFKQEMKMTKIKK-LEKETTYRWSWSSNKALLEMA 431
DB 1827 ARQLRLEDELNRAKSTLEAEITRVKQRLCEKQIQNDLQNMKTQYSRKEBAIRK--IESE 1884
QY 432 EKTVRDK-----ELEGVQKIQRLKLCR-----ALQTERNDLNKRVQDLSA 474
DB 1885 REKSERKNSLRSEIERLQAEIKRIEERCCKRLKEDSTRETOSQLETYSRYQREIDKL-- 1942
QY 475 GGQSLTDSGPERPEGPGCAQAPSPVTEAPC 507
DB 1943 -----RQRPYG-----SHRETQTEC 1957

RESULT 34

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
XX Claim 20; Page 266-267; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAW79323-AAW80102) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3656 (AAW80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
XX Sequence 931 AA;
XX SQ

```

Query Match          9.4%; Score 254.5; DB 4; Length 931;
Best local Similarity 21.4%; Pred. No. 2.le-07;
Matches 104; Conservative 99; Mismatches 170; Indels 113; Gaps 14;

QY 33  SQAPRKPEGAQARTAQSCALRDV---SEZLSRQLEDILSTCYVDNVOGGPDCGAQGEPA 89
Db 371  SEAPQKSGHGVNELTSRLUKLEMNQSLTKTVEELRTTV-----DSVEGNAS 419
QY 90  EPEDAEKSRVTVAANGSPPEPTPVYGE-----KPSKQDPNTEERIQSDEV 135
Db 420  KILQVKEHQSLSKVVEILENIIVQEKSQLCNQNLKDLMKEXAQLEKTIEITRENSE- 478
QY 136  GDRHRRAPQSKKAGKGLKEITILMQTINTLS-----TPBEKLAALCKKVAELLEHRNS 190
Db 479  -----RQIKILEQENHEHLNQTVSSLRQSRQISAEARVQDIEKNKILHESIKET 527
QY 191  QKQMKLLQKQSQSLVQEKDHLRGHXSNAVLAERSKLSLCELOHNN----- 236
Db 528  SSKLSKLEFRKQKKELEHYKEGEKA-----SELENEHLHLEKENLLQKKITNLKITC 583
QY 237  ---RSLKEEGVQRAREBEERKEKVTSHFQVTLNDIQLMQEOHNERNKSLROENNELAERL 293
Db 584  EKTEALEQENSELERENKUKKTILDS-----FNKLTQLESLEKENSQLODENLELRNV 638
QY 294  KKLIEQVELREEHDKVYFKHDLQOQLVDKLAQQAQEMLKAEERHOREKDFLLKEA--- 350
Db 639  ESL-----KCAISMQAQLOLENKESEKEQLKKGLE-----LLKASFKK 678
QY 351  -----VESQRMCBLKQOEHTLQOOLALYTEKEPEFPQNTLSKSEVEFTTQOE 398
Db 679  TERLEVSYQGLDENQBLQKTLNENKKI--QOLESBIQDLEMENQTTQKQLELKISKR 737
QY 399  MEKMTKKIKLEKETTMY---RSRWSSNKALLEMAEBEKTVRDKEGLEGVQKIORLEKLC 455
Db 738  LEQULEKNKSLEQETSQLEKDKQKLEKNKELRQAE---IKOTTLENNVKNIGLEKEN 794
QY 456  RALQTE 461
Db 795  KTLKSKE 800

```

```

RESULT 36
AAW78520
ID AAW78520 standard; protein; 990 AA.
XX
XX
XX AAW78520;
XX
XX
XX 06-NOV-2001 (first entry)
XX
XX Human protein SEQ ID NO 1182.
DE

```

KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;
KW	nervous system disorder; arthritis; inflammation.
OS	Homo sapiens.
XX	
WO	200157190-A2.
XX	
09-AUG-2001.	
XX	
05-FEB-2001; 2001WO-0004098.	
XX	
03-FEB-2000; 2000US-00496914.	
PR	27-APR-2000; 2000US-00560875.
PR	20-JUN-2000; 2000US-00598075.
PR	19-JUL-2000; 2000US-00620325.
PR	01-SEP-2000; 2000US-00654936.
PR	15-SEP-2000; 2000US-00663561.
PR	20-OCT-2000; 2000US-00693325.
PR	30-NOV-2000; 2000US-00728422.
XX	
(HYSE-) HYSEQ INC.	
PA	
XX	
Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;	
PI	Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren P, Chen R, Wang ZW;
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX	
WPI; 2001-476283/51.	
DR	N-PSDB; AAK51453.
XX	
Nucleic acids encoding polypeptides with cytokine-like activities, useful	
PT	in diagnosis and gene therapy.
XX	
Claim 20; Page 3425-3427; 6221pp; English.	
XX	
The invention relates to polynucleotides (AAK51456-AAK53435) and the	
CC	encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to
CC	cytokine, cell proliferation or cell differentiation or which may induce
CC	production of other cytokines in other cell populations. The
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC	peptide therapy. The polypeptides have various cytokine-like activities,
CC	e.g. stem cell growth factor activity, haematopoiesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	activin/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC	inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC	(AAK52582) and 3466 (AAM80020) are omitted as the relevant pages from the
CC	sequence listing were missing at the time of publication

```

XX SQ Sequence 990 AA;
Query Match          9.4%; Score 254.5; DB 4; Length 990;
Best Local Similarity 21.4%; Pred. No. 2.2e-07;
Matches 104; Conservative 99; Mismatches 170; Indels 113; Gaps 14;

QY      33   SQAPRPEGAQAATAQSALRV--SELSROLEILSTYCVDNNGGPGEDGAQEPA    89
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Dd      430 SEAPQSLGHVENLTISRLLKLEMNQSLTKTVBELRTV-----DSVEGNAS  478

QY      90 EPEDAQRTRYVARNEGEPTPVYGE-----KEPSKDPPNTESIROSDEV    135
       ::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Dd     479 KILMKEKENQLRSKKVEILEINEIVQEKSQLONGQNLSKOLMKEAQLKETIETLRNSE-  537

QY     136 GDDRHRPQBCKAKGLKGELITLMLMTATLS---TPBEKLAAALCKCYAEILLBHRNS  190
       :||||:||||:||||:||||:||||:||||:||||:||||:||||:
Dd     538 -----RQIKILEQENHNLNTVSLSQRSQISAZARVKDIKENKILHESIKET    586

QY     191 QOKMKLLQKQSQUVOBKHLRGHSKAVALARSKLESBLERPHN-----        236
       ::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Dd     587 SSKLKIEFEKQJIKKBELHYEYXGBRA----EELENELHFHLEKENELLQKITNKUIC  642

QY     237 ---RLSKEGVGRAREEBBKRKEVTSHFOTVLNDILOMQEHQHNRNSKLROENNELAERL  293

```

Db	643	EKTEAEQSENERNRKLTLDLSD-----FKNLTFQLESLEKNSQDSENLERNV	697	CC	disease. (M1) is useful for identifying a compound which modulates the
Qy	294	KXIEQVELREHIDKFKHKLQQLVDKLAQQAQEMKAEERHOREKDFLLKEA---	350	CC	activity of a gene product, preferably enzymatic activity, carbon
Db	698	ESL-----KCSMKMAQLQLENKELSESEKQKGLG-----LLKASPKK	737	CC	compound catabolism, biosynthetic, transporter, transcriptional,
Qy	351	-----VESQRMCLMKQOETHLQOALYTEKPEEPQNTLSKSSEVFTTFKQE	398	CC	translational, signal transduction, DNA replication and cell division
Db	738	TERLEVSVOGLDIENORLQNTLSNKKI-QQLESELQDLEMNQTLQKNLELKISSKR	796	CC	activity. The method is useful for identifying a compound having the
Qy	399	MEKWTXIKKLEKETWY---RSRWSSNKALLEMAEETKVRDKELEGQVQTKQLRKLK	455	CC	ability to inhibit growth or proliferation of C. albicans cells and for
Db	797	LEQLEKNSLQETSQLEKDKQLEKENVRLQQAQLE---IKDTTLENNVKLGNLEKEN	853	CC	treating infection by C. albicans. The present sequence is that of an
Qy	456	RALQTE 461		CC	essential Candida albicans protein used in the method of the invention.
Db	854	KTLSE 859		CC	Note: The sequence data for this patent is not represented in the printed
				CC	specification but is based on sequence information supplied to Derwent by
				XX	the European Patent Office
				Seq	Sequence 1881 AA;
					Query Match 9.4%; Score 253.5; DB 5; Length 1881;
					Best Local Similarity 21.6%; Pred. No. 5.4e-07;
					Matches 100; Conservative 94; Mismatches 195; Indels 73; Gaps 12;
Qy	52	LRDVSEELSRQLEDJLST-----YCVNNQGGPGDGAQGEPAEPDAEKSRTRYVARNGP	107	Db	776 LEDKEQLAKIQEDHKSLNEKELVTANSLOGIKATKSESTISGPDQQLQALQKGNIS
Db	776	LEDKEQLAKIQEDHKSLNEKELVTANSLOGIKATKSESTISGPDQQLQALQKGNIS	835	Qy	108 EPTPVVYGEKPSKGPNTETIRQSDVGDGRDHR-----PQEKKKAKG
Qy	108	EPTPVVYGEKPSKGPNTETIRQSDVGDGRDHR-----PQEKKKAKG	151	Db	836 ESTLKQKLEKLS-----TEQAKKLEDGJINNMTDLPFLKSKSGAEATQIKQREPEKN
Db	836	ESTLKQKLEKLS-----TEQAKKLEDGJINNMTDLPFLKSKSGAEATQIKQREPEKN	890	Qy	152 L-----GKEITLMTLNTLSTP-EKLAALCKKYAELLEHEHNSQKMK-----
Qy	152	L-----GKEITLMTLNTLSTP-EKLAALCKKYAELLEHEHNSQKMK-----	195	Db	891 LTFEFTKQYELQINNLINKSNBPKINELSKISLTDNFKPNKQLEKLDTEB
Db	891	LTFEFTKQYELQINNLINKSNBPKINELSKISLTDNFKPNKQLEKLDTEB	950	Qy	196 ----LLQKQSQOLVQEKDHLRGE---HSAVLARSKLSESLCRLQHNHNSLKEEGVQAR
Qy	196	----LLQKQSQOLVQEKDHLRGE---HSAVLARSKLSESLCRLQHNHNSLKEEGVQAR	248	Db	951 NNEHLMDLRSASVAYNDLKAKASEEETVAKKELETLTLSKIDNLEKELKEQ-QSKKN
Db	951	NNEHLMDLRSASVAYNDLKAKASEEETVAKKELETLTLSKIDNLEKELKEQ-QSKKN	1009	Qy	249 EEEKKEKVTSHFQVTLNDIQLOMEQHNERNSKLQENMELAEKRLKLEQVELREHID
Qy	249	EEEKKEKVTSHFQVTLNDIQLOMEQHNERNSKLQENMELAEKRLKLEQVELREHID	308	Db	1010 ELEGQNTITDSTNEKFKLEDELKSIKSNSELSQSELSQLEKTEKQLEKQAKDEID
Db	1010	ELEGQNTITDSTNEKFKLEDELKSIKSNSELSQSELSQLEKTEKQLEKQAKDEID	1069	Qy	309 KYFKEKDLQOQLVDKLAQQAQEMKAEERHOREKDFLLKEAVESORMCMLKQOETHLK
Qy	309	KYFKEKDLQOQLVDKLAQQAQEMKAEERHOREKDFLLKEAVESORMCMLKQOETHLK	368	Db	1070 KLAETKSNIDNINSEISSLSQSLKAESEHSHTVD---EHSLSLENLKKLEEVNTKT
Db	1070	KLAETKSNIDNINSEISSLSQSLKAESEHSHTVD---EHSLSLENLKKLEEVNTKT	1126	Qy	369 QQLALYTEKPEEPQNTLSKSSEVFTTFKQEMERMTKIKKLEKETTMYRSRWESSNKAL
Qy	369	QQLALYTEKPEEPQNTLSKSSEVFTTFKQEMERMTKIKKLEKETTMYRSRWESSNKAL	428	Db	1127 SMIAKLSAKIEBHKA-----TDEIETTKTHITDLOBEHAKQSQFESRNDIK
Db	1127	SMIAKLSAKIEBHKA-----TDEIETTKTHITDLOBEHAKQSQFESRNDIK	1175	Qy	429 ENAEKTVRDKELEGQVQIKQLEKLCRALQTERNDLANKRVQ 470
Qy	429	ENAEKTVRDKELEGQVQIKQLEKLCRALQTERNDLANKRVQ 470		Db	1176 SNLDE---ANKELSDNRKLSNLEK-----EXTELANKLK 1207
Db	1176	SNLDE---ANKELSDNRKLSNLEK-----EXTELANKLK 1207			
				RESULT 38	
				ADD46507	
				ID	ADD46507 standard; protein; 1206 AA.
				XX	AC ADD46507;
				XX	DT 29-JAN-2004 (first entry)
				XX	DE Rat Protein BAA20077, SEQ ID NO 12188.
				XX	Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
				XX	Chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
				XX	Rattus norvegicus.
				OS	WO2003016475-A2.
				XX	27-FEB-2003.
				PD	14-AUG-2002; 2002WO-US025765.
				XX	14-AUG-2001; 2001US-0312147P.
				PR	

CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. Gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1206 AA;

Query Match 9.4%; Score 253; DB 7; Length 1206;
Best Local Similarity 22.2%; Pred. No. 3.4e-07;
Matches 133; Conservative 92; Mismatches 199; Indels 176; Gaps 25;

QY 12 EGAQEPSPQAAPAVEAGPSSOA-----PR 37
DB 682 ENAGELPVKAEP-----QAPAAQSPPPVLPISNIHSENTKGMALPKPETILPP 737
QY 38 KPFGAQAARTAQSG---ALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGPAPEDA 94
DB 738 EPENGKNDTSGTGVNENSSDNLISFSL-----KTKDSGVSILQETRRQKTL 791
QY 95 EKRTYVARGPEPTFVYVGEKPKGDNTEIRQSDVEGDRHRRPQEKKAQGLQK 154
DB 792 KTKRKFIVDGVESVTT---SKIVTSDSKTEELR-----FLRQELFELLLQK 838
QY 155 EITLLMOTLN-TLSTPEEKLAALCKYAEILLBHRNSQOMKLLQKQSQQLVQEKDHLRG 213
DB 839 ESKAQOOLNGKLOQREQIFRPEQ---EMLSKRGYDQBIENLEKQKQTH----- 889
QY 214 EHSKAVLARSLSRELQHRNSLKBEQVORAREEERKKEVTSHFQVTLNDIQLQWE 273
DB 890 -----RLS-----QHTNRLRDE-AKRIKGEQKE---LSKFQNLN----- 923
QY 274 QHNRNSKLQENMELAEELKLEO-----YELREH 306
DB 924 RKKEBEFVQKQOQLDGLAKLIIQOKAEIANIERCLNKKOQLARARAEALWELEERH 983
QY 307 IDKVFHKDIQQQLVDKLAQQAQEMLEKAE---ERHQREKDFLLKEAV---ESQRMCELMK 361
DB 984 LQE--KHQLLKQQLKQDYFQFHQLLKREHETEQQVRYNORLI BELKNRQTOQERARLPK 1041
QY 362 QQTHLKQQLALY-----TEKFEFQNTLSKSSEVTTTFKQEMKATKKI 406
DB 1042 IQRSEAKTRMAFKKSLRINSTATPDQDREKIKQFAAQEEK-----RQKNERMAQH- 1092
QY 407 KLEKETTYVRSWSSNKKALLEMAEK-----TVRDKLE---GLQVKIQRLKLC 455
DB 1093 QKHESQWRDLQACEANVRELHQIQNEKCHLLVEHETQKLDEHESQELKEWR-EKLR 1151
QY 456 RALQTERNDLNKRVQDLS-----AGGGSLTDSGPRRPPGGAQAPSPRVTEAPCPYG 510
DB 1152 PRKKTLEEPARKLQEQVEVFQMTGSECLNPSAQSR-----GCLQTSHPSSSTRAPAWAG 1206

RESULT 40
AAB96332
ID AAB96332 standard; protein; 880 AA.
XX AC AAB96332;
XX DT 29-OCT-2001 (first entry)
XX DE Putative P. abyssi ATPase involved in DNA repair #2.
XX KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX OS Pyrococcus abyssi.
XX PR 292651-A1.

XX 27-OCT-2000.

PD 21-APR-1999; 99FR-00005034.

XX 21-APR-1999; 99FR-00005034.

XX (CNRS) CNRS CENT NAT RECH SCI.

PA (IPRE-) IPREMER INST FR RECH EXPL MER.

XX Porterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

PI Querellou J, Weissenbach J, Saurin W, Hellig R;

XX WPI; 2001-126236/14.

DR New nucleotide sequences isolated from Pyrococcus abyssi encode proteins

PT useful in industry.

XX Claim 7; Page 1003-1006; 1657pp; French.

XX The present invention relates to the genomic sequence of Pyrococcus

CC abyssi (see AAF86431 and AAF41223-7) and P. abyssi proteins. P. abyssi is

CC a hyperthermophilic archaeon, which is isolated from deep-sea

CC hydrothermal vents. The present sequence is one such P. abyssi protein.

CC The proteins of the present invention have various potential industrial

CC uses, since the proteins are stable at very high temperatures, some up to

CC 110 degrees centigrade. Note: This patent is in the same patent family as

CC WO200065062, which contains additional sequences as shown in AAB99132-

CC AAB99143, AAB75903-AAH75920 and AAG66436

XX Sequence 880 AA;

QY 12 EGAQEPSPQAAPAVEAGPSSQ-----APRKEGAQAARTAQSGALRDVSEELSRQLE 64

DB 259 KGLEEKIYQIERSIEBKAKISELBEIVKDIPLQEKKEKYLKXG-FRDEYESKLRLLE 317

QY 65 DILSTYCVNNQGGPGEDGAQGPAPEDAQKSRTRYVARGPEPTFVYVGEKPKSGDP 124

DB 318 KELSFW-----ESLKAIBEVIKEG-----KKKB 342

QY 125 NTEEIRQS-DEVGDR-DHRRP--QEKKAKGLGKBITLLMOTLNTLSTPE--EKLAAALCK 178

DB 343 RAEEIREKLSERKLEBELKPTVELEDAKQVQKQIERLKARLKLSPGEVTEKLESLEK 402

QY 179 KYAEELLEHRNSQOMKLLQKQSQQLVQEKDHLRGEHSAVLARSKLSLC-----REL-Q 233

DB 403 ERTEI-----EAAKEITTRIGOMEQKN-----ERKKAIEELKAKGKCPVCGRLETE 451

QY 234 RHNRSLK-----EGVQARAEERKEKKEVTSHFQVTLNDIQLQ---MEQHNRNS 280

DB 452 EHKELMERYTLIEIKIIEELKRTTTEBRKRLVNRKLEIKLUREFSVMDIAQIKLEBS 511

QY 281 KLROENMELAE-----RLKKGLEIQLYELREHIDKVPK 312

DB 512 KLKGFNLEELQKEREPEGLNEEFNKLKGLGLERDLKRIKALSGRRKLIBEYKVKACE 571

QY 313 H-KLQOQQLVDAKLQQAQEM---LKAEEHOREKDFLLKEAVESQRMCELMKQETHLK 368

DB 572 ELENLHRLQRLGFPESVEELNLRIQELEEFHDKYVAAKXSES-ELRELKMLKLEKTELD 630

QY 369 QQLALYTBKFEFQNTLSKSSEVFTTF--KQEMKATKKIKLEKETTYVRSWSSNKA- 426

DB 631 QAFEMLADVENIEIEKEAKLDLESKFNEEYEEKERLKVLEREVSSILARLEELKKS 690

QY 427 -----LLEMAEKTVRDKBLEGLQVKTORLEKLCRALQTERNDLNKRVQD 471

DB 691 EQKATLAKLKEEKEREK---AKLEIKKLEKALSKE---DLRKKIKD 733

```
RESULT 41
ABB61144
ID ABB61144 standard; protein; 1690 AA.
XX
XX
AC ABB61144;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 10224.
XX
XX Drosophila melanogaster polypeptide; cell signalling; insecticide;
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL05247.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 10224; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1690 AA;
XX
XX Query Match 9.3%; Score 251.5; DB 4; Length 1690;
XX Best Local Similarity 23.3%; Pred. NO. 6.3e-07;
XX Matches 127; Conservative 102; Mismatches 180; Indels 135; Gaps 23;
XX
XX 33 SQAPKPGGAQRTAQSGALDVSELSRQLEDILSTYCVUNQCGFGDGAQGEPAEPE 92
XX 352 SSTPVKPLATPKSQFS--MODLREKQOQVKEKLMVERDLDR-----E 392
XX
XX 93 DAERSRTYVARNGEPEPTPVYGEKPSKDPNTEERQS--DEV---GDRDHRPQ-EXK 147
XX
XX 393 DAQWALQLOXNINELKXRIVELESALDNERRKTEELQCSIDEAQFCGDELNAQSQVYKE 452
XX
XX 148 KAKGKGKEITLMO---TINTLSTP-----EKLALCKK-----YALL 184
XX
XX 453 KIHDLSEKITLVNATPSLQSLPDLPSDDGALQEEIAKLQEKVTIOQKEVESRIAEOL 512
XX
XX 185 EERNRSOKMLQK---OSQVQEKDHLRGEHSAVLARSKLCSLRELQRHNSLK 240
XX
XX 513 EEQRLRNKVLNQTALQSLVSKDEAL-----EKFSLSECCINLRLE---LLX 564
XX
XX 241 EGVQRABEE-----EKGKEV-----TSHFOVTLNDI 268
XX
XX 565 BENEKQAQAEFTKLAERKSEVILRLSELQNLKATSDSLSERVKNKTDECEILQTEV 624
```

```

269 OLQMBQHNRNSKL-----RQENNELAE-----RLKKLIEQYELREHIDKVFKHDL 316
625 RMRDEQIRELNQOLDEVTTQLVNQKADSSALDDMLRLQK--EGTEKSTLLKKT----- 676
317 QQQLYDAKLQQAQEMLKAEERHQRKDFLLKAEAVESQRMCELMKQ-----QETHLK 368
677 EKELVQSK-EQAATLNDKQLEKQISD--LKQLAEQKLVREMTENALNQIQLEKESIE 733
369 QQALALYTERKFEFQNTLSKSSEVFTTFKQ-----EMEKMTKKIKKLEKKTMTYSRWE 421
734 QQALAKQNELEDFOKKQSSSEVHLQEIKAQNTQKQFELVESGESLKKLQQLQEQKTLGHS 793
422 SSNKALLENABEKTIV-----RDELEGLQVKIQLEKLCALQTERNDLNKRVQDLSAGQ 477
794 KLQALAEELKKEKKEKTIIEKEQELOQLOSKSABESAEKVVQVQLEQLQO--QAASGEE 851
478 GSLT 481
852 GSKT 855

RESULT 42
ABB61173
ID ABB61173 standard; protein; 1690 AA.
XX
XX ABB61173;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 10311.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL05276.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 10311; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1690 AA;
XX
XX Query Match 9.3%; Score 251.5; DB 4; Length 1690;
XX
```


XX	18-FEB-2000 (first entry)	Db	168	MREGDVTMSNSVHLKPEENYREGDP-----RT---RASDPSPQSVSRHKS	215
XX	Human SULU3 protein.	Qy	119	PSKGPNTBEIRQSDVGRDRHRRPQKCKAKGLCKEIT-----LLMQTLNTLTSTPBKLA	174
XX		Db	216	HYRNRHFPATIRTSALV-----TRQMEHQDSLAQMSGYKRRRQHQKQLATLENKIK	271
XX		Qy	175	ALCKYA-----ELLEHRN--SOKWKLQKQSOINVQ-----KDLRGEHSHK	217
XX		Db	272	AEWDEHRLDKOLETQORNFAPAEKLIKQQAAMEKAKVMSKEKFKQHQIQQQKK	331
XX		Qy	218	AVLARSKLBSLCRLQRNRSIKBEQVQ-----ARBEHE-----KRKEVTSHPQV--TLNDI	268
XX		Db	332	EL--NSFLSQKREYKLRKEQLKEELNENQSTPKKEQSWLSKQKQENIQHQAEBEANLL	389
XX		Qy	269	QLOME-----CHN-----ERNKLRQENMELAEKRLKLIQVELRE	304
XX		Db	390	RHQROVLELCRRFKRMMLGRHNLQDLVRELKNRQTKQDLHMLLRQSHSQBLEF	449
XX		Qy	305	EHDVKVFKHDKLQQQLVDKLAQAQEMLKAERHQENDFLKKAESQRCMLKQOE	364
XX		Db	450	RHLNTIQK---MRCELL--RLQHOTELTNQLEYNKRRERELRRKRVMEVRCQPSLSKSE	504
XX		Qy	365	THLKQQL-----ALYTKFEEFQNTLSKSEVTTTQKQEMKMTKKIKLEKTTMYRSW	420
XX		Db	505	LQIKQFQDTCKIQTRQYKALRNHLE-----TTPKSEHKAVLKRLK--EEQTKRLAILA	557
XX		Qy	421	BESNKALLEMAEBKTVR-----DKELE	442
XX		Db	558	EQYDSINEMLSTQALRLDEAQAECQVLKWLQLELILNAYQSKIQKQASQAQHDREL	617
XX		Qy	443	GLQVKI-----QRLEKLCRALQTERND-----LNKRVQDLASAGQGS-----L	480
XX		Db	618	ELEQVSLRRALLEQKIEEMALQNERTERIRSLERQAREAFDSMSRLGFSMMVL	677
XX		Qy	481	TDSGPERREPGCAQAPSPRVTEAPCYGAP--STEASGQTGP	522
XX		Db	678	SNLSPE-----AFSHSYPGASGWNPTGGQGP	705
XX	RESULT 45				
XX	AAV55942				
XX	ID	AAV55942		standard; peptide; 1001 AA.	
XX	AC	AAV55942;			
XX	DT	19-FEB-2000 (first entry)			
XX	DE	Human/Murine SULU3 consensus protein sequence.			
XX	KW	Antirheumatic; antiarthritic; antiinflammatory; antiallergic; osteopathic;			
XX	KW	antipsoriatic; antiarteriosclerotic; antiasthmatic; immunosuppressive;			
XX	KW	neuroprotective; cardiant; cerebroprotective; cytostatic; antidiabetic;			
XX	KW	vulnery; STE20; protein kinase; STLK2; STLK3; STLK4; STLK5; STLK6; STLK7;			
XX	KW	ZC1; ZC2; ZC3; ZC4; KHS2; SULU3; SULU3; GSK2; PAK4; PAK5; antagonist;			
XX	KW	antibody; gene therapy; rheumatoid arthritis; artherosclerosis; asthma;			
XX	KW	inflammatory bowel disease; Crohn's disease; osteoarthritis; psoriasis;			
XX	KW	rhinitis; autoimmunity; organ transplantation; multiple sclerosis;			
XX	KW	myocardial infarction; cardiovascular disease; stroke; renal failure;			
XX	KW	oxidative stress-related neurodegenerative disorder; Parkinson's disease;			
XX	KW	amyotrophic lateral sclerosis; Leigh syndrome; cancer; cardiomyopathy;			
XX	KW	ischemic disorder; inflammation; diabetes mellitus; fibrosis; mitosis;			
XX	KW	mesangial disorder; growth regulation; wound healing; T cell activation;			
XX	KW	immunosuppressant.			
XX	OS	Homo sapiens.			
XX	OS	Mus sp.			
XX	XX	MO9953036-A2.			
XX	XX	21-OCT-1999.			
XX	XX	13-APR-1999;		99WO-US008150.	

Antirheumatic; antiarthritic; antiinflammatory; antiallergic; osteopathic; antipsoriatic; antiarteriosclerotic; antiasthmatic; immunosuppressive; neuroprotective; cardiant; cerebroprotective; cytostatic; antidiabetic; vulnery; STE20; protein kinase; STLK2; STLK3; STLK4; STLK5; STLK6; STLK7; ZC1; ZC2; ZC3; ZC4; KHS2; SULU3; SULU3; GSK2; PAK4; PAK5; antagonist; antibody; gene therapy; rheumatoid arthritis; artherosclerosis; asthma; inflammatory bowel disease; Crohn's disease; osteoarthritis; psoriasis; rhinitis; autoimmunity; organ transplantation; multiple sclerosis; myocardial infarction; cardiovascular disease; stroke; renal failure; oxidative stress-related neurodegenerative disorder; Parkinson's disease; amyotrophic lateral sclerosis; Leigh syndrome; cancer; cardiomyopathy; ischemic disorder; inflammation; diabetes mellitus; fibrosis; mitosis; mesangial disorder; growth regulation; wound healing; T cell activation; immunosuppressant.

Homo sapiens.

MO9953036-A2.

21-OCT-1999.

13-APR-1999; 99WO-US008150.

14-APR-1998; 98US-0081784P.

(SUGE-) SUGEN INC.

Plowman G, Martinez R, Whyte D;

WPI; 1999-611301/52.

N-PSDB; RAZ40489.

Novel kinase-related polypeptides used for the diagnosis and treatment of kinase-related diseases and disorders.

Claim 11; Page 296-299; 387pp; English.

This sequence represents a novel STE20-related protein kinase. The invention relates to nucleic acid molecule encoding a kinase polypeptide selected from STLK2, STLK3, STLK4, STLK5, STLK6, STLK7, ZC1, ZC2, ZC3, ZC4, KHS2, SULU1, SULU3, GSK2, PAK4 and PAK5. The proteins are used to identify agonists and antagonists, and to raise antibodies. The polynucleotides are useful in gene therapy protocols. The polynucleotides, antibodies, antagonists and agonists may be used to treat diseases such as immune-related disorders and diseases (e.g. rheumatoid arthritis, artherosclerosis, chronic inflammatory bowel disease (e.g. Crohn's disease), asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplantation, chronic inflammatory pelvic disease, multiple sclerosis, organ transplantation, myocardial infarction, cardiovascular disease, stroke, renal failure, oxidative stress-related neurodegenerative disorders (e.g. amyotrophic lateral sclerosis, Parkinson's disease and Leigh syndrome), cancer, cardiomyopathies, ischemic disorders, inflammatory disorders, diabetes mellitus, fibrotic and mesangial disorders. The proteins may also be useful for cell growth regulation (e.g. in wound healing), T cell activation, mitosis control, and as immunosuppressants

Sequence 786 AA;

Query Match 9.2%; Score 248.5; DB 2; Length 786;

Best Local Similarity 21.6%; Pred. No. 4e-07;

Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

QY 23 PAVEAE-----GFGSSQA--PRKEGAQAFTAQSGALRVSELSQLE 64

109 PAVEAQEEERQDHGVRGTGVNSVGSNQSTPSMSISASSQSSSVNSLPDVSDDKS-ELD 167

QY 65 DILSTYCVNNQG-----GFGEDG---AQGEPAEPEDAESKRTYVANGEPPTPVVYGEK 118

XX 14-APR-1998; 98US-0081784P.
XX (SUGS-) SUGEN INC.
XX Plowman G, Martinez R, Whyte D;
XX WPI; 1999-611301/52.
XX Novel kinase-related polypeptides used for the diagnosis and treatment of
XX kinase-related diseases and disorders.
XX Claim 11; Page 312-315; 387pp; English.
XX This sequence represents a consensus peptide sequence contained in novel
XX STE20-related protein kinases. The invention relates to a nucleic acid
XX molecule encoding a kinase polypeptide selected from STIK2, STIK3, STIK4,
XX STIK5, STIK6, STIK7, ZC1, ZC2, ZC3, ZC4, KHS2, SULU3, GEX2, PAK4
XX and PAK5. The proteins are used to identify agonists and antagonists, and
XX to raise antibodies. The polynucleotides are useful in gene therapy
XX protocols. The polynucleotides, polypeptides, antibodies, antagonists and
XX agonists may be used to treat diseases such as immune-related disorders
XX and diseases (e.g. rheumatoid arthritis, arteriosclerosis, chronic
XX inflammatory bowel disease (e.g. Crohn's disease), asthma,
XX osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and
XX organ transplantation, chronic inflammatory pelvic disease, multiple
XX sclerosis, organ transplantation, myocardial infarction, cardiovascular
XX disease, stroke, renal failure, oxidative stress-related
XX neurodegenerative disorders (e.g. amyotrophic lateral sclerosis,
XX Parkinson's disease and Leigh syndrome), cancer, cardiomyopathies,
XX ischemic disorders, inflammatory disorders, diabetes mellitus, fibrotic
XX and mesangial disorders. The proteins may also be useful for cell growth
XX regulation (e.g. in wound healing), T cell activation, mitosis control,
XX and as immunosuppressants
XX
XX Sequence 1001 AA;
Query Match 9.2%; Score 248.5; DB 2; Length 1001;
Best Local Similarity 21.6%; Pred. No. 5.3e-07;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;
QY 23 PAVEAB-----GPGSSQA-PRKPEGAQARTAQSGALRDYSELSROLE 64
DB 324 PAVEAQEEBEEBQHGVRTGTVNSVNSQSPMSISASSQSSVNSLPDVSDDKS-ELD 382
QY 65 DILSTYCVNNQ-----GPGEDG--AQGEPAEPDAEKSRITYVARNGEPEPTVWYGEKE 118
DB 383 MMEGDHVMNSSVIHLKPEENVREGDP-----RT---RASDPQSPQVSRHKS 430
QY 119 PSKGDPTNTEIROSDVGRDHRPQRKKAAGLGEIT-----ILMOTLNTLSTPEKLA 174
DB 431 HYRNREHFATIRTSALV-----TRQMGEHQDSLELREQMSGYKMRMRQHOKQLMTLENK 486
QY 175 ALCKKVA-----ELLEERN--SOKOMKLLQKDSOLVOE-----KDLRGEHKS 217
DB 487 AEMDEHLRLDKDLETGRNFAEMKELIKKHAQAEKEAKVMSNEKKFQHIQAOQK 546
QY 218 AVIARSKLESCLRELQHNLSLKEGVQR-----AREEEE--KKRVTGHFQV--TINDI 268
DB 547 EL--NSFLSQKREYKLRQLKELNENQSTPKCKQEWLSKQENIQHFAEEBANLL 604
QY 269 QLOWE-----QNN-----ERNKLRQENNELAERLKLKLEQYELAE 304
DB 605 RRORVYLESCRPKRMLGRNLEODLVRELNKRQTOKOLEHAMLRLQHESMGLEFP 664
QY 305 EHKDKVPMKDLQOOLVDKALQQAENLKEABERHOREKDFLKEAVESQRMCELMKQOE 364
DB 665 RLHNTIQK---MRCELI--RLQHTQLTNTQLEYNKRERELRKRHVMYVRQPKSLKSX 719
QY 365 THLKOQL-----ALYTEKFEFQNTLSKSEVFTTFQEMEKMTKKLKLKETTMTYRSW 420
DB 720 LOIKQFQDCTCKIQTRQYKALRNHLE-----TTPKSEHKAVLKLK--BEQTKRLAILA 772

QY 421 ESNKALLEMAEKTVR-----DKLELE 442
DB 773 EYVDHSINEMLSQALRLDEAQEAECQVLKMQOLQQLLELLNAYQSKIKMQAEQHDRELR 832
QY 443 GLQVKI-----QLEKLCEALOTERND-----LNKRVODLSAGGQS-L 480
DB 833 ELQQRVSLRRALLEQKIEEMALQKQERIRSLRQAREIEAFDSMRGLGFSNMVL 892
QY 481 TDSGPERPEPGCAQAPSSPRVTEAPCYGAP--STEASGOTGP 522
DB 893 SNLSPE-----AFSHSYDGAAGWSHNPFGGPGP 920
RESULT 46
AB897326
ID ABB97326 standard; protein; 1001 AA.
XX
XX ABB97326;
XX
XX 27-JUN-2002 (first entry)
XX Novel human protein SEQ ID NO: 594.
XX Human; antianaemic; vulnary; antiinflammatory; immunomodulator;
XX antifertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
XX neuroprotective; antiparkinsonian; protein therapy; EST;
XX expressed sequence tag.
XX Homo sapiens.
XX WO200222660-A2.
XX 21-MAR-2002.
XX 10-SEP-2001; 2001WO-US026015.
XX 11-SEP-2000; 2000US-00659671.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-292408/33.
XX N-PSDB; ABN32512.
XX An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis.
XX Example 2; SEQ ID NO 594; 509pp; English.
XX The present invention provides the protein and coding sequences of 444
XX novel human proteins. These are isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
XX and cancer, to screen for drugs, to treat inflammatory conditions e.g.
XX rheumatoid arthritis, and to treat nervous system disorders e.g.
XX Parkinson's disease. The present sequence is a protein of the invention
XX
XX Sequence 1001 AA;
Query Match 9.2%; Score 248.5; DB 5; Length 1001;
Best Local Similarity 21.6%; Pred. No. 5.3e-07;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;
QY 23 PAVEAB-----GPGSSQA-PRKPEGAQARTAQSGALRDYSELSROLE 64
DB 324 PAVEAQEEBEEBQHGVRTGTVNSVNSQSPMSISASSQSSVNSLPDVSDDKS-ELD 382
QY 65 DILSTYCVNNQ-----GPGEDG--AQGEPAEPDAEKSRITYVARNGEPEPTVWYGEKE 118

Db 773 EQYDHSINELMSTQALRLDEAQAECQVLKQQLQQLLELLNAYQSKIKWQAQAQHDREL 832
QY 443 GLQVKI-----QLEKLCRALQTERND-----LNKRVODLSAGQGS-----L 480
Db 833 ELEQVSLRLALAEQTEEMALQNERTERIRSLERQARETEAFDSMRLGFSNMVL 892
QY 481 TDSGPERPBGQAQAPSSPRVTEAPCYGAP--STEASGQTGP 522
Db 893 SNLSPE-----AFSHSYFGASGWSHNPTGGPGP 920

RESULT 48
ABR47509
ID ABR47509 standard; protein; 1005 AA.
AC ABR47509;
XX
DT 12-JUN-2003 (first entry)
DB Breast cancer associated protein sequence SEQ ID NO:254.
KW Human; breast cancer; cytostatic; gene therapy.
XX Homo sapiens.
XX WO2003004989-A2.
XX
XX 16-JAN-2003.
XX
XX 21-JUN-2002; 2002WO-US019669.
XX
XX 21-JUN-2001; 2001US-0299887P.
XX 27-JUN-2001; 2001US-0301572P.
XX 18-JUL-2001; 2001US-0306501P.
XX 25-SEP-2001; 2001US-0325002P.
XX 05-MAR-2002; 2002US-0362585P.
XX 14-MAY-2002; 2002US-0380391P.
XX
XX (MILL-) MILLENIUM PHARM INC.
XX
XX Lillie J, Gannavarapu M, Giatt K, Hoersh S, Kamatkar S;
XX Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;
XX Bast RC, Hortobagyi GN, Puzstai L, Meric P, Sahin A, Mills GB;
XX
XX WPI; 2003-210381/20.
XX N-PSDS; ACC50205.
XX
XX Breast cancer diagnosis or treatment by comparing the level of expression
XX of a marker in a patient sample with that in the control non-breast
XX cancer sample.
XX
XX Claim 1; SEQ ID NO 254; 128pp; English.
XX
XX The present invention describes a method for assessing whether a patient
XX is afflicted with breast cancer. The method comprises comparing the level
XX of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
XX ABR47386 to ABR47632) in a patient sample and the normal level of
XX expression of the marker in a control non-breast cancer sample, where a
XX significant increase in the level of expression of the marker in the
XX patient sample and the normal level is an indication that the patient is
XX afflicted with breast cancer. The breast cancer associated sequences from
XX the present invention have cytostatic activities and can be used in gene
XX therapy. The method is useful for diagnosing and treating breast cancer.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences

Sequence 1005 AA;
Query Match 9.2%; Score 248.5; DB 6; Length 1005;
Best Local Similarity 21.6%; Pred. No. 5.3e-07;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

QY 23 PAVTAE-----GPGSQA-PRKPEGQAARTQAQSGALRDVSELSROLE 64
Db 328 PAVTAEQEEBEOHVGRTGTVNSVGSQSLIPSMISISSQSSVNSLPDVSDDKS-ELD 386
QY 65 DILSTYCYDNNQG-----GPGEDG--AQGEPAEPDAEKSRITYVAINGRPEPTPVYGRKE 118
Db 387 XMEGDHTVMSNSSVTHLKPEENYREGDP-----RT---RASDQSPQVSRHKS 434
QY 119 PSKGPNTETEQSDVGDROHRRPOEKKKAKGLKEIT---LLMQTLNTLTSTPEEKL 174
Db 435 HYRNRHFPATIRTSLV---TRQCEHEQDSLEQMSGYRMRROHQQLMTLENLKL 490
QY 175 ALCKKYA-----ELLEHERN--SOKQMLQKQSQOLVOS-----KDLRGEHSK 217
Db 491 AEMDEHRLDKDLEFORNPAEEMKLIKXQAAMEKAQVMSNEEKFOQHIOAQQKK 550
QY 218 AVLARKSLSLCRLQHRNRSLSKEGVQR-----AREESE---KRKEVTSHFQV--TLNDI 268
Db 551 EL--NSFLSSQKREYKLRKEQLKEBLNENQSTPKKEKQEWLSKQENIQHFOREANLL 608
QY 269 QLQME-----QHN-----ERNSKLRQENNELAERIKKLEQVYELRE 304
Db 609 RQRCQVLELCRRFKRBMMLGRHNLEQDIVREELNKRQTKDLEHAWLLRQESMQELBF 668
QY 305 EHDVKVFKHQLQOLVDAKLQQAQEMLKBAERHOREKDFLLKEAVESQRCMLMKQOE 364
Db 669 RHLNTIQK---NRCELL--RLQFQTELTNQLSYNKRERERELRRKHVMBVROQPKSLKSKE 723
QY 365 THLKQOL-----ALYTEKPEEFQNTLSKSSEVFTFKQEMKMKTKIKKLEKETTYRSRW 420
Db 724 LQIKQFPQDTCKIQTOYKALRNHLL-----TTPSKHKA VLKRLK--EETRKLAILA 776
QY 421 ESSNKALLMMAEKTVR-----DKELE 442
Db 777 EQYDHSINELMSTQALRLDEAQAECQVLKQQLQQLLELLNAYQSKIKWQAQAQHDREL 836
QY 443 GLQVKI-----QLEKLCRALQTERND-----LNKRVODLSAGQGS-----L 480
Db 837 ELEQVSLRLALAEQTEEMALQNERTERIRSLERQARETEAFDSMRLGFSNMVL 896
QY 481 TDSGPERPBGQAQAPSSPRVTEAPCYGAP--STEASGQTGP 522
Db 897 SNLSPE-----AFSHSYFGASGWSHNPTGGPGP 924

RESULT 49
AAO30953
ID AAO30953 standard; protein; 1005 AA.
XX
XX AAO30953;
XX
XX 06-OCT-2003 (first entry)
XX Human TAOJIX protein #3 used to modulate beta-catenin pathway.
XX
XX Human; beta-catenin; TAOJIX; gene therapy; endometrial cancer; breast;
XX bladder; stomach stromal sarcoma; cytostatic.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Domain 32..285
XX /note= "Kinase domain"
XX
XX WO2003051905-A2.
XX
XX 26-JUN-2003.
XX
XX 12-DEC-2002; 2002WO-US039742.
XX
XX 13-DEC-2001; 2001US-0340312P.
XX

PA	(EXEL-) EXELIXIS INC.	
XX	Costa MA, Gendreau SB, Dora EG, Nicoll M, Urbani L, Larson JS;	
XX	WPI; 2003-514040/48.	
XX	Identifying a candidate beta-catenin pathway modulating agent for	
PT	diagnosing or treating cancer by detecting a test agent-biased activity	
PT	of the assay system comprising a purified TAOJIK polypeptide or nucleic	
PT	acid.	
XX		
PS	Claim 13; Page 88-92; 92pp; English.	
XX		
CC	The invention relates to a method for identifying a candidate beta-	
CC	catenin pathway modulating agent. The method comprises providing an assay	
CC	system comprising a purified TAOJIK polypeptide or nucleic acid or its	
CC	functionally active fragment or derivative; contacting the assay system	
CC	with a test agent under conditions where the system provides a reference	
CC	activity except for the presence of the test agent; and detecting a test	
CC	agent-biased activity of the assay system; where a difference between the	
CC	test agent-biased activity and the reference activity identifies the	
CC	agent as a candidate beta-catenin pathway modulating agent. The method is	
CC	useful in gene therapy and for manufacturing a medicament for diagnosing	
CC	or treating breast, bladder or endometrial cancer or stomach stromal	
CC	sarcoma. The present sequence is human TAOJIK protein used to modulate	
CC	beta-catenin pathway	
XX		
SQ	Sequence 1005 AA;	
	Query Match 9.2%; Score 248.5; DB 6; Length 1005;	
	Best Local Similarity 21.6%; Pred No. 5.3e-07;	
	Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;	
QY	23 PAVAE-----GPGSSQA-PRKEGAQARTAQSGALRDVSELSQLE 64	
DB		
DB	328 PAVAEQEEEDHGCGRTGTWNSVGSNSQIPSMISIGASSQSSVNSLPDVSDKKS-ELD 386	
QY	65 DILSTYCVNNQG-----GPGEDG--ACGEPAEPEDAESRTYVARNGEPEPTPVYGEKE 118	
DB		
DB	387 MMEGDHTVMSNSVTHLKPEENYRBECDP-----RT---RASDPQSPPOVSRHKS 434	
QY	119 PSKGPNTTEIRQSDVEGDRDRPQEKKAQKGLKEIT-----LLMOTLNTLSTPEKLA 174	
DB		
DB	435 HYRREHPATINTASLV---TRQMEHQDSSELREQMSGYKMRQHQKQMLTENLKLK 490	
QY	175 ALCKKYA---ELLEHRN--SQOMKLLQKQSOVLQV-----KQHLRGHNSK 217	
DB		
DB	491 AEMDEHRLRLDKOLETQRNNFAAEMEKLIKXQAAEMKEBAKVMGNEKKFOOHTQAOQKX 550	
QY	218 AVLARSKLESLELQRLHNRSLKEEGVQV---ARREEE---KKEVTSHPQV---TLNDI 268	
DB		
DB	551 EL--NSFLESQKREYKLRKEQLKHELNEHQSTPKKEQKQWLSKQENIQHQAEEBANLL 608	
QY	269 QLOVE-----QHN-----ERNKLRQENMELAEKLLKLIQYELRE 304	
DB		
DB	609 RRQOYLELCRRFRKRRMLLGRHNLQDLVRELNKRTQDKLEHAMLRRQHSQMLELF 668	
QY	305 EHIDKVPKHLQOOLVDKLAQAEKLEABERHOREKDFLLKEAVESQKMLKQOE 364	
DB		
DB	669 RHLNTIQK--WRCELI--RLQHOTELTNQLEYKRRERELRRHVRVHQPKSLKSKE 723	
QY	365 THLKOQL-----ALYTEKPEFQNTLSKSSVFYTFKQEMKWKYKKLEKETTYRSRW 420	
DB		
DB	724 LQIKKQFOQTCKIQTROYKALRNHLE-----TTPKSHKAVLKLK--BEQTKLAILA 776	
QY	421 ESSNKALLMAEAKTVR-----DKELE 442	
DB		
DB	777 EYDYSINEMLSQTALRDEDAQEAECQVLKMQLOQELBNLNAQSKIKMQABAOHDELR 836	
QY	443 GLOVKI-----QRLKLCRALQTERND-----LNKRVODLSAGQOGS-----L 480	
DB		
DB	837 ELEQVSLRRALLEQKIEEMALQNTERTIRSLRQAREIEAFDESRLGFSNNVL 896	

QY

481

TDSGPERPEGCAQAPSSPRVTEAPCYGAP--STEASGQTGP 522

DB

897

SNLSP-----APSHSYFGASGSHNPTGGPGP 924

RESULT 50

ABP42281

ID

ABP42281 standard; protein; 85 AA.

XX

AC

ABP42281;

DT

22-AUG-2002 (first entry)

XX

Human ovarian antigen HITJAS0, SEQ ID NO:3413.

XX

Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;

XX

Human cancer; breast cancer; tumour; reproductive system disorder;

XX

infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;

XX

PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;

XX

inflammatory condition; immune disorder; blood disorder;

XX

cardiovascular disorder; respiratory disorder; neurological disorder;

XX

gastrointestinal disorder; urinary system disorder; drug screening;

XX

gene therapy; chromosome mapping; forensic analysis;

XX

antibody preparation; cytostatic; immunomodulatory; neuroprotective;

XX

antiinflammatory; gynaecological; reproductive.

XX

Homo sapiens.

XX

WC20020677-A1.

XX

03-JAN-2002.

XX

07-JUN-2001; 2001WO-US018569.

XX

07-JUN-2000; 2000US-0209467P.

XX

(HUMA-) HUMAN GENOME SCI INC.

XX

Birse CE, Rosen CA;

XX

WPI; 2002-147878/19.

DR

N-PSDB; ABQ55358.

XX

Isolated nucleic acid molecules encoding novel ovarian polypeptides,

XX

useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian

XX

cancer), immune disorders, cardiovascular disorders and neurological

XX

diseases.

XX

Claim 11; SEQ ID NO 3413; 2922pp; English.

XX

The invention relates to 2175 novel human ovarian antigens (ABP41054-

XX

ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also

XX

encompasses polypeptides 90% identical and polynucleotides 95% identical

XX

to the sequences of the invention. The invention additionally relates to

XX

recombinant vectors and host cells comprising human ovarian antigen

XX

polynucleotides, antibodies against human ovarian antigens, and the use

XX

of ovarian antigen polynucleotides and polypeptides in diagnosing,

XX

treating, prognosing or preventing various ovary and/or breast-related

XX

disorders. Such conditions include ovarian cancer and breast cancer, and

XX

metastatic tumours of ovarian or breast origin, reproductive system

XX

disorders (e.g., infertility, disorders of pregnancy, anovulation,

XX

polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine

XX

disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic

XX

shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and

XX

vaginitis), immune disorders (e.g., congenital and acquired

XX

immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),

XX

blood-related disorders (e.g., anaemia), cardiovascular disorders,

XX

respiratory disorders, neurological disorders, gastrointestinal disorders

XX

and urinary system disorders. Ovarian antigen polypeptides and

XX

polynucleotides may also be used in screening for compounds which

XX

modulate ovarian antigen expression or activity. The polynucleotides may

XX

further be used for gene therapy, chromosome mapping, in the

XX

identification of individuals and in forensic analysis, and the

CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 85 AA;

Query Match 9.1%; Score 247; DB 5; Length 85;
Best Local Similarity 64.5%; Pred. No. 3.7e-08;
Matches 49; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

QY 347 LKEAVESORMCELMKQOETHLKQQLALYTEKPEEFQNTLSKSSEVFTTPEKQEMKMTKKI 406
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2 LKEATESRHKYBOMKQXEVQLKQQLSLYMGKPEEFQNTWAKSNELVTTPEKQEMKMTKKI 61

QY 407 KLEKETTMYRSRWES 422
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
62 KLEKETTMYRSRWES 77

Search completed: June 7, 2004, 14:37:32
Job time : 149 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:37:41 ; Search time 127 Seconds
(without alignments)
1174.091 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 2702

Sequence: 1 KSSPGQPEAGPEGAQERPSQ.....APSTEASGQTGPQETPSAZA 530

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2702	100.0	530	9	US-09-962-055-8
2	2702	100.0	530	9	US-09-976-740-8
3	2702	100.0	530	12	US-10-671-242-8
4	2702	100.0	530	13	US-10-023-529-8
5	2702	100.0	530	13	US-10-023-523-8
6	2702	100.0	530	15	US-10-616-187-8
7	2693	99.7	546	9	US-09-976-740-44
8	2693	99.7	546	12	US-10-671-242-44
9	2693	99.7	546	13	US-10-023-529-44
10	2693	99.7	546	13	US-10-023-523-44
11	2693	99.7	546	15	US-10-616-187-44
12	2465.5	91.2	557	9	US-09-962-055-5
13	2465.5	91.2	557	9	US-09-976-740-5
14	2465.5	91.2	557	12	US-10-671-242-5
15	2465.5	91.2	557	13	US-10-023-529-5

16	2465.5	91.2	557	13	US-10-023-523-5	Sequence 5, Appli
17	2465.5	91.2	557	15	US-10-616-187-5	Sequence 5, Appli
18	2344	86.0	510	12	US-10-276-774-2134	Sequence 2134, Ap
19	501	18.5	437	12	US-10-424-599-207434	Sequence 55188, A
20	468	17.3	436	12	US-10-425-114-55188	Sequence 207434,
21	444.5	16.5	314	12	US-10-425-114-43684	Sequence 43684, A
22	375	13.9	346	15	US-10-369-493-5130	Sequence 5130, Ap
23	304.5	11.3	206	12	US-10-425-114-45419	Sequence 45419, A
24	259	9.6	886	15	US-10-369-493-1016	Sequence 1016, Ap
25	257.5	9.5	860	12	US-10-072-012-838	Sequence 838, App
26	257.5	9.5	860	12	US-10-037-417-59	Sequence 59, Appli
27	257.5	9.5	860	15	US-10-080-334-166	Sequence 166, App
28	256	9.5	2871	14	US-10-146-473-41	Sequence 41, Appli
29	253.5	9.4	1881	14	US-10-032-585-7646	Sequence 7646, Ap
30	252.5	9.3	880	15	US-10-369-493-21643	Sequence 21643, A
31	248.5	9.2	786	10	US-09-291-417-23	Sequence 23, Appli
32	248.5	9.2	1001	10	US-09-291-417-31	Sequence 31, Appli
33	248.5	9.2	1005	14	US-10-177-233-254	Sequence 254, App
34	248.5	9.2	1005	14	US-10-317-835-18	Sequence 18, Appli
35	247	9.1	85	15	US-10-264-049-3413	Sequence 3413, Ap
36	247	9.1	1001	15	US-10-445-735-2	Sequence 2, Appli
37	244.5	9.0	909	9	US-09-925-299-988	Sequence 988, App
38	244.5	9.0	909	10	US-09-925-299-988	Sequence 1, Appli
39	242	9.0	1534	14	US-10-342-136-1	Sequence 102, App
40	238	8.8	892	14	US-10-205-219-102	Sequence 1168, Ap
41	237	8.8	1938	16	US-10-408-765A-1168	Sequence 16, Appli
42	236.5	8.8	1137	12	US-10-336-472-16	Sequence 254, App
43	236.5	8.8	3225	16	US-10-408-765A-254	Sequence 1629, Ap
44	236	8.7	1938	16	US-10-408-765A-2188	Sequence 2188, Ap
45	235.5	8.7	1939	16	US-10-408-765A-2188	Sequence 3636, Ap
46	234.5	8.7	650	15	US-10-104-047-3636	Sequence 20, Appli
47	234	8.7	1859	12	US-10-336-472-20	Sequence 22, Appli
48	234	8.7	1859	12	US-10-336-472-22	Sequence 4080, Ap
49	233.5	8.6	512	15	US-10-108-260A-4080	Sequence 11, Appli
50	230.5	8.5	3878	14	US-10-080-608A-11	Sequence 2, Appli
51	230.5	8.5	3899	14	US-10-171-311-4	Sequence 4, Appli
52	230.5	8.5	3907	14	US-10-171-311-2	Sequence 100, App
53	230.5	8.5	3811	15	US-10-370-685-100	Sequence 1839, Ap
54	230.5	8.5	3911	16	US-10-408-765A-1839	Sequence 8, Appli
55	230.5	8.5	3917	14	US-10-171-311-8	Sequence 6, Appli
56	230.5	8.5	3925	14	US-10-171-311-6	Sequence 96, Appli
57	228.5	8.5	1941	12	US-10-188-186-96	Sequence 564, Ap
58	228	8.4	1164	15	US-10-369-493-6564	Sequence 260, App
59	228	8.4	2383	14	US-10-082-830-260	Sequence 47959, A
60	227.5	8.4	660	9	US-09-864-761-47959	Sequence 2653, Ap
61	227.5	8.4	678	15	US-10-264-049-2653	Sequence 2324, A
62	227.5	8.4	820	14	US-10-029-386-2324	Sequence 7776, Ap
63	227.5	8.4	1038	14	US-10-032-585-7776	Sequence 2, Appli
64	227.5	8.4	1162	12	US-09-894-273-2	Sequence 2, Appli
65	227.5	8.4	1162	14	US-10-294-804-2	Sequence 73, Appli
66	227	8.4	888	10	US-09-893-519A-73	Sequence 1928, Ap
67	226.5	8.4	698	15	US-10-094-749-1328	Sequence 2119, Ap
68	226.5	8.4	1114	16	US-10-408-765A-2119	Sequence 1586, Ap
69	226.5	8.4	1790	15	US-10-369-493-1586	Sequence 99, Appli
70	226.5	8.4	1940	10	US-09-738-630-99	Sequence 107, App
71	226	8.4	1999	15	US-10-028-248A-107	Sequence 107, App
72	226	8.4	1999	15	US-10-107-782-107	Sequence 10206, A
73	224	8.3	1180	15	US-10-369-493-10206	Sequence 105, App
74	224	8.3	1961	15	US-10-028-248A-105	Sequence 105, App
75	224	8.3	1961	15	US-10-107-782-105	Sequence 63, Appli
76	223.5	8.3	670	12	US-09-298-523B-63	Sequence 63, Appli
77	223.5	8.3	670	12	US-09-748-875-63	Sequence 831, App
78	223.5	8.3	735	12	US-10-087-192-831	Sequence 19, Appli
79	223.5	8.3	868	9	US-09-884-001-19	Sequence 36, Appli
80	223.5	8.3	1959	15	US-10-028-248A-36	Sequence 36, Appli
81	223.5	8.3	1959	15	US-10-107-782-36	Sequence 5, Appli
82	223	8.3	1379	14	US-10-205-219-5	Sequence 5, Appli
83	223	8.3	1974	15	US-10-369-493-6395	Sequence 6395, Ap
84	222.5	8.2	701	10	US-09-298-523B-62	Sequence 62, Appli
85	222.5	8.2	701	12	US-09-748-875-62	Sequence 1175, Ap
86	222.5	8.2	1940	16	US-10-408-765A-1175	Sequence 449, App
87	221.5	8.2	1043	15	US-10-310-154-449	Sequence 6328, Ap
88	221	8.2	1091	15	US-10-369-493-6328	

89 221 8.2 1857 16 US-10-408-765A-2173 Sequence 2173, Ap
90 221 8.2 1961 15 US-10-028-248A-103 Sequence 103, App
91 221 8.2 1961 15 US-10-107-782-103 Sequence 103, App
92 220.5 8.2 1742 15 US-10-012-697-1548 Sequence 1548, Ap
93 219.5 8.1 690 10 US-09-298-523B-61 Sequence 61, Appl
94 219.5 8.1 690 12 US-09-748-875-61 Sequence 61, Appl
95 219.5 8.1 2035 15 US-10-080-334-160 Sequence 160, App
96 219 8.1 1123 14 US-10-342-136-3 Sequence 3, Appl
97 218.5 8.1 623 15 US-10-104-047-3378 Sequence 3378, Ap
98 218.5 8.1 1203 14 US-10-097-340-43 Sequence 43, Appl
99 218 8.1 443 8 US-08-325-278-6 Sequence 6, Appl
100 218 8.1 1017 16 US-10-408-765A-158 Sequence 158, App
101 218 8.1 1156 15 US-10-369-493-43 Sequence 43, Appl
102 218 8.1 3336 16 US-10-408-765A-2453 Sequence 2453, Ap
103 217.5 8.0 967 14 US-10-080-608A-21 Sequence 21, Appl
104 217.5 8.0 967 15 US-10-370-685-110 Sequence 110, App
105 217.5 8.0 1956 15 US-10-369-493-6729 Sequence 6729, Ap
106 217 8.0 621 15 US-10-108-260A-4409 Sequence 4409, Ap
107 217 8.0 707 10 US-09-298-523B-2 Sequence 2, Appl
108 217 8.0 707 12 US-09-748-875-2 Sequence 2, Appl
109 217 8.0 2041 16 US-10-408-765A-1151 Sequence 1151, Ap
110 217 8.0 2099 14 US-10-128-714-3290 Sequence 3290, Ap
111 217 8.0 2405 14 US-10-128-714-8290 Sequence 8290, Ap
112 216.5 8.0 903 12 US-10-282-122A-52328 Sequence 52328, A
113 216.5 8.0 1001 14 US-10-128-714-3240 Sequence 3240, Ap
114 216.5 8.0 1384 16 US-10-473-576-22 Sequence 22, Appl
115 216.5 8.0 1404 16 US-10-473-576-2 Sequence 2, Appl
116 216 8.0 711 10 US-09-298-523B-3 Sequence 3, Appl
117 216 8.0 711 12 US-09-748-875-3 Sequence 3, Appl
118 216 8.0 715 16 US-10-408-765A-2096 Sequence 2096, Ap
119 216 8.0 762 15 US-10-334-143-32 Sequence 32, Appl
120 216 8.0 1583 16 US-10-408-765A-1635 Sequence 1635, Ap
121 215.5 8.0 1938 14 US-10-171-311-164 Sequence 164, App
122 215.5 8.0 1945 15 US-09-927-597-2 Sequence 2, Appl
123 215.5 8.0 1959 15 US-10-028-248A-106 Sequence 106, App
124 215.5 8.0 1959 15 US-10-107-782-106 Sequence 106, App
125 215.5 8.0 1960 15 US-10-236-031B-62 Sequence 62, Appl
126 215.5 8.0 1960 15 US-10-028-248A-104 Sequence 104, App
127 215.5 8.0 1960 15 US-10-107-782-104 Sequence 104, App
128 215.5 8.0 1972 14 US-10-171-311-162 Sequence 162, App
129 215.5 8.0 1972 15 US-10-341-434-103 Sequence 103, App
130 215.5 8.0 1979 10 US-09-927-597-4 Sequence 4, Appl
131 215 8.0 758 16 US-10-408-765A-687 Sequence 687, App
132 215 8.0 1837 15 US-10-369-493-22734 Sequence 22734, A
133 214.5 7.9 466 15 US-10-108-260A-4344 Sequence 4344, Ap
134 214.5 7.9 1031 11 US-09-764-875-686 Sequence 686, App
135 214.5 7.9 1035 15 US-10-158-057-197 Sequence 197, App
136 214.5 7.9 2033 16 US-10-408-765A-419 Sequence 419, App
137 214 7.9 550 15 US-10-104-047-3495 Sequence 3495, Ap
138 214 7.9 1300 16 US-10-408-765A-257 Sequence 257, App
139 214 7.9 1302 16 US-10-408-765A-1995 Sequence 1995, Ap
140 214 7.9 1957 15 US-10-369-493-2070 Sequence 2070, Ap
141 213.5 7.9 1978 15 US-10-094-466-64 Sequence 64, Appl
142 213 7.9 691 10 US-09-298-523B-1 Sequence 1, Appl
143 213 7.9 691 12 US-09-748-875-1 Sequence 1, Appl
144 213 7.9 733 12 US-10-424-599-167699 Sequence 167699, App
145 213 7.9 929 10 US-09-298-523B-60 Sequence 60, Appl
146 213 7.9 929 12 US-09-748-875-60 Sequence 60, Appl
147 213 7.9 929 16 US-10-299-636-94 Sequence 94, Appl
148 213 7.9 933 15 US-10-369-493-12831 Sequence 12831, A
149 213 7.9 1294 12 US-10-282-122A-61292 Sequence 61292, A
150 213 7.9 1596 15 US-10-080-334-42 Sequence 42, Appl

ALIGNMENTS

RESULT 1
US-09-962-055-8
; Sequence 8, Application US/09562055
; Patent No. US2002052033A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.

Lees, Robert S.
Law, Simon W.
Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSES: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/962,055
FILING DATE: 24-Sep-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/979,608
FILING DATE: 26-NOV-1997
APPLICATION NUMBER: US 60/031,930
FILING DATE: 27-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 1983/59818)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-962-055-8

Query Match 100.0%; Score 2702; DB 9; Length 530;
Best Local Similarity 100.0%; Fred. No. 3.3e-156; Indels 0; Gaps 0;
Matches 530; Conservative 0; Mismatches 0;
QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGFGSSQAPRKEGAQARTAQSGALRDVSEELS 60
DB 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGFGSSQAPRKEGAQARTAQSGALRDVSEELS 60
QY 61 RQLEDILSTYCYDNNQGGFGEDGAQGEPAEPEDAKSRFYVARNGEPEPTPVVYGEKPS 120
DB 61 RQLEDILSTYCYDNNQGGFGEDGAQGEPAEPEDAKSRFYVARNGEPEPTPVVYGEKPS 120
QY 121 KGPNTPEERQSDVEGDRDRHPQEKKAAGLKGKXITLLMQTLNTLSTPEEKLAALCKKY 180
DB 121 KGPNTPEERQSDVEGDRDRHPQEKKAAGLKGKXITLLMQTLNTLSTPEEKLAALCKKY 180
QY 181 AEILLEHRNSQOMKLLQKKQSQLVQEKDLHGEHSKAVLARSKLESICELORHRSK 240
DB 181 AEILLEHRNSQOMKLLQKKQSQLVQEKDLHGEHSKAVLARSKLESICELORHRSK 240
QY 241 ERGVQAREEERKEKVTSHFQVTLNDIOLQMEQHNRNSKLQRQENNELAERLKKLIEQY 300
DB 241 ERGVQAREEERKEKVTSHFQVTLNDIOLQMEQHNRNSKLQRQENNELAERLKKLIEQY 300
QY 301 ELRREHIDKVFKEKDLQOQLVDAKQQAQEMLKEAFERHOREKDFLLKKAVERSORCELM 360
DB 301 ELRREHIDKVFKEKDLQOQLVDAKQQAQEMLKEAFERHOREKDFLLKKAVERSORCELM 360
QY 361 KQOETHLKOALALYTEKFEFQNTLKSSEVFTTFQEMEKTKTKLKEKETTWYRSRW 420

Db 361 KQOETHLKQOLALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
Qy 421 ESSNKALLEMAEETKVRDKELEGQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLS 480
Db 421 ESSNKALLEMAEETKVRDKELEGQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLS 480
Qy 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTASGOTGPEPTSARA 530
Db 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTASGOTGPEPTSARA 530

RESULT 2
US-09-976-740-8
; Sequence 8, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-8

Query Match 100.0%; Score 2702; DB 9; Length 530;
Best Local Similarity 100.0%; Pred. No. 3.3e-156;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGFGSSQAAPKPEGAQARTAQSGALRDVSELS 60
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGFGSSQAAPKPEGAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVNNQGGPGEAGGAPPAVEAEGFGSSQAAPKPEGAQARTAQSGALRDVSELS 120
Db 61 RQLEDILSTYCVNNQGGPGEAGGAPPAVEAEGFGSSQAAPKPEGAQARTAQSGALRDVSELS 120
Qy 121 KGDPTTEIRQSDVGDHRDHRPQEKKAAGLGEITLLMQTLNTLSTPEKLAALCKKY 180
Db 121 KGDPTTEIRQSDVGDHRDHRPQEKKAAGLGEITLLMQTLNTLSTPEKLAALCKKY 180
Qy 181 AELLEHRNSQKMKLLQKQSQQLVQKDHLSGHSKAVLARSKLESCLRELQHRNSLK 240
Db 181 AELLEHRNSQKMKLLQKQSQQLVQKDHLSGHSKAVLARSKLESCLRELQHRNSLK 240
Qy 241 BEGVQARAREEKEKKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENNELAERLKKLIEQY 300
Db 241 BEGVQARAREEKEKKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENNELAERLKKLIEQY 300
Qy 301 ELREEHIDKVFPHKDIQOQLVDKLAQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
Db 301 ELREEHIDKVFPHKDIQOQLVDKLAQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
Qy 361 KQOETHLKQOLALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
Db 361 KQOETHLKQOLALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420

Qy 421 ESSNKALLEMAEETKVRDKELEGQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLS 480
Db 421 ESSNKALLEMAEETKVRDKELEGQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSLS 480
Qy 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTASGOTGPEPTSARA 530
Db 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTASGOTGPEPTSARA 530

RESULT 3
US-10-671-242-8
; Sequence 8, Application US/10671242
; Publication No. US20040040049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; CURRENT FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-8

Query Match 100.0%; Score 2702; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 3.3e-156;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGFGSSQAAPKPEGAQARTAQSGALRDVSELS 60
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGFGSSQAAPKPEGAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVNNQGGPGEAGGAPPAVEAEGFGSSQAAPKPEGAQARTAQSGALRDVSELS 120
Db 61 RQLEDILSTYCVNNQGGPGEAGGAPPAVEAEGFGSSQAAPKPEGAQARTAQSGALRDVSELS 120
Qy 121 KGDPTTEIRQSDVGDHRDHRPQEKKAAGLGEITLLMQTLNTLSTPEKLAALCKKY 180
Db 121 KGDPTTEIRQSDVGDHRDHRPQEKKAAGLGEITLLMQTLNTLSTPEKLAALCKKY 180
Qy 181 AELLEHRNSQKMKLLQKQSQQLVQKDHLSGHSKAVLARSKLESCLRELQHRNSLK 240
Db 181 AELLEHRNSQKMKLLQKQSQQLVQKDHLSGHSKAVLARSKLESCLRELQHRNSLK 240
Qy 241 BEGVQARAREEKEKKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENNELAERLKKLIEQY 300
Db 241 BEGVQARAREEKEKKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENNELAERLKKLIEQY 300
Qy 301 ELREEHIDKVFPHKDIQOQLVDKLAQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
Db 301 ELREEHIDKVFPHKDIQOQLVDKLAQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
Qy 361 KQOETHLKQOLALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
Db 361 KQOETHLKQOLALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420

Db 361 KQETHLKKQALALYTERFEBFQNTLSKSSSEVFTTFKQEMENMTKKIKLEKETTYWRSW 420
Qy 421 ESSNKALLEWAEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVQDLNKGQGS 480
Db 421 ESSNKALLEWAEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVQDLNKGQGS 480
Qy 481 TDSGPERRPEGCAQAPSSPRVTAPCPYCAPSTASGOTGPQPTPSARA 530
Db 481 TDSGPERRPEGCAQAPSSPRVTAPCPYCAPSTASGOTGPQPTPSARA 530

RESULT 4

US-10-023-529-8
; Sequence 8, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-529-8

Query Match 100.0%; Score 2702; DB 13; Length 530;
Best Local Similarity 100.0%; Pred. No. 3.3e-156;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGCAQARTAQSGALRDVSELS 60
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGCAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVDNNOGGPGDGAQGEPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Db 61 RQLEDILSTYCVDNNOGGPGDGAQGEPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Qy 121 KGDPTNTEIROSDEVDGRDHRPQEKKAAGLGKEITLLMOTLNTLSTPEKLAALCKKY 180
Db 121 KGDPTNTEIROSDEVDGRDHRPQEKKAAGLGKEITLLMOTLNTLSTPEKLAALCKKY 180
Qy 181 AEILLEHRSNOKMKLLQKKSQVQKDLRGHESKAVLARSKLESICRELQRHNSLK 240
Db 181 AEILLEHRSNOKMKLLQKKSQVQKDLRGHESKAVLARSKLESICRELQRHNSLK 240
Qy 241 EGVQARAEERKEKRVTSHFQVTLNDIQLQMEQHNEHNSKLRQENNELAERLKKLIEQY 300
Db 241 EGVQARAEERKEKRVTSHFQVTLNDIQLQMEQHNEHNSKLRQENNELAERLKKLIEQY 300
Qy 301 ELREHHIDKVPFKHDLQOOLVDKLOQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 360
Db 301 ELREHHIDKVPFKHDLQOOLVDKLOQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 360
Qy 361 KQETHLKKQALALYTERFEBFQNTLSKSSSEVFTTFKQEMENMTKKIKLEKETTYWRSW 420

Db 361 KQETHLKKQALALYTERFEBFQNTLSKSSSEVFTTFKQEMENMTKKIKLEKETTYWRSW 420
Qy 421 ESSNKALLEWAEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVQDLNKGQGS 480
Db 421 ESSNKALLEWAEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVQDLNKGQGS 480
Qy 481 TDSGPERRPEGCAQAPSSPRVTAPCPYCAPSTASGOTGPQPTPSARA 530
Db 481 TDSGPERRPEGCAQAPSSPRVTAPCPYCAPSTASGOTGPQPTPSARA 530

RESULT 5

US-10-023-523-8
; Sequence 8, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-523-8

Query Match 100.0%; Score 2702; DB 13; Length 530;
Best Local Similarity 100.0%; Pred. No. 3.3e-156;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGCAQARTAQSGALRDVSELS 60
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGCAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVDNNOGGPGDGAQGEPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Db 61 RQLEDILSTYCVDNNOGGPGDGAQGEPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Qy 121 KGDPTNTEIROSDEVDGRDHRPQEKKAAGLGKEITLLMOTLNTLSTPEKLAALCKKY 180
Db 121 KGDPTNTEIROSDEVDGRDHRPQEKKAAGLGKEITLLMOTLNTLSTPEKLAALCKKY 180
Qy 181 AEILLEHRSNOKMKLLQKKSQVQKDLRGHESKAVLARSKLESICRELQRHNSLK 240
Db 181 AEILLEHRSNOKMKLLQKKSQVQKDLRGHESKAVLARSKLESICRELQRHNSLK 240
Qy 241 EGVQARAEERKEKRVTSHFQVTLNDIQLQMEQHNEHNSKLRQENNELAERLKKLIEQY 300
Db 241 EGVQARAEERKEKRVTSHFQVTLNDIQLQMEQHNEHNSKLRQENNELAERLKKLIEQY 300
Qy 301 ELREHHIDKVPFKHDLQOOLVDKLOQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 360
Db 301 ELREHHIDKVPFKHDLQOOLVDKLOQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 360

Qy 361 KQETHLKQALALYTERKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTYRSRW 420
Db 361 KQETHLKQALALYTERKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTYRSRW 420
Qy 421 ESSNKALLEMAEKTVDKLEGLQVQIKQRIORLEKLCRALQTERNDLNKRVQDLSAGGGS 480
Db 421 ESSNKALLEMAEKTVDKLEGLQVQIKQRIORLEKLCRALQTERNDLNKRVQDLSAGGGS 480
Qy 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPCAPSTEASGQTGPQPTTSARA 530
Db 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPCAPSTEASGQTGPQPTTSARA 530

RESULT 6

US-10-616-187-8
; Sequence 8, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-8

Query Match 100.0%; Score 2702; DB 15; Length 530;
Best Local Similarity 100.0%; Pred. No. 3.3e-156;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKPGEAQAARTAQSGALRDVSEELS 60
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKPGEAQAARTAQSGALRDVSEELS 60
Qy 61 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPEDAERKSTYVARNGBEPPTPVVYGEKPS 120
Db 61 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPEDAERKSTYVARNGBEPPTPVVYGEKPS 120
Qy 121 KGPNTBEIRQSDVEGDRDRRPPQKKAGLGKEITLLMOTLNTLTSTPEKLAALCKKY 180
Db 121 KGPNTBEIRQSDVEGDRDRRPPQKKAGLGKEITLLMOTLNTLTSTPEKLAALCKKY 180
Qy 181 AELLFEHRNSQKMKLLQKQSQVLQVEKDHRLRGEHSKAVLARSKLESICRELQHRNRSK 240
Db 181 AELLFEHRNSQKMKLLQKQSQVLQVEKDHRLRGEHSKAVLARSKLESICRELQHRNRSK 240
Qy 241 EGVQARAEERKEKVTSHFQVTLNDIQLOMEQHNRNSKLRQENNELAERLKLIEQY 300
Db 241 EGVQARAEERKEKVTSHFQVTLNDIQLOMEQHNRNSKLRQENNELAERLKLIEQY 300
Qy 301 ELREEHIDKVPFKHKLQOQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRCMLM 360
Db 301 ELREEHIDKVPFKHKLQOQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRCMLM 360

Qy 361 KQETHLKQALALYTERKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTYRSRW 420
Db 361 KQETHLKQALALYTERKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTYRSRW 420
Qy 421 ESSNKALLEMAEKTVDKLEGLQVQIKQRIORLEKLCRALQTERNDLNKRVQDLSAGGGS 480
Db 421 ESSNKALLEMAEKTVDKLEGLQVQIKQRIORLEKLCRALQTERNDLNKRVQDLSAGGGS 480
Qy 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPCAPSTEASGQTGPQPTTSARA 530
Db 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPCAPSTEASGQTGPQPTTSARA 530

RESULT 7

US-09-976-740-44
; Sequence 44, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-44

Query Match 99.7%; Score 2693; DB 9; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.2e-155;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKPGEAQAARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKPGEAQAARTAQSGALRDVSEELS 76
Qy 61 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPEDAERKSTYVARNGBEPPTPVVYGEKPS 120
Db 77 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPEDAERKSTYVARNGBEPPTPVVYGEKPS 136
Qy 121 KGPNTBEIRQSDVEGDRDRRPPQKKAGLGKEITLLMOTLNTLTSTPEKLAALCKKY 180
Db 137 KGPNTBEIRQSDVEGDRDRRPPQKKAGLGKEITLLMOTLNTLTSTPEKLAALCKKY 196
Qy 181 AELLFEHRNSQKMKLLQKQSQVLQVEKDHRLRGEHSKAVLARSKLESICRELQHRNRSK 240
Db 197 AELLFEHRNSQKMKLLQKQSQVLQVEKDHRLRGEHSKAVLARSKLESICRELQHRNRSK 256
Qy 241 EGVQARAEERKEKVTSHFQVTLNDIQLOMEQHNRNSKLRQENNELAERLKLIEQY 300
Db 257 EGVQARAEERKEKVTSHFQVTLNDIQLOMEQHNRNSKLRQENNELAERLKLIEQY 316
Qy 301 ELREEHIDKVPFKHKLQOQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRCMLM 360
Db 317 ELREEHIDKVPFKHKLQOQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRCMLM 376

Qy 361 KQETHLKQALALYTEKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKLEKETTMTSRW 420
Db 377 KQETHLKQALALYTEKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKLEKETTMTSRW 436
Qy 421 ESSNKALLEWAEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQSSL 480
Db 437 ESSNKALLEWAEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQSSL 496
Qy 481 TDSGPERRPEGPAQAPSSPRVTAPCYPGAPSTEASGQTGPQPTTSARA 530
Db 497 TDSGPERRPEGPAQAPSSPRVTAPCYPGAPSTEASGQTGPQPTTSARA 546

RESULT 8
US-10-671-242-44
; Sequence 44, Application US/10671242
; Publication No. US20040040049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; CURRENT FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-44

Query Match 99.7%; Score 2693; DB 12; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.2e-155;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 76
Qy 61 RQLEDILSTYCVDNNGGPGEDGAGGPAEPEDAESKRTYVARGPEPTPVVYGEKPS 120
Db 77 RQLEDILSTYCVDNNGGPGEDGAGGPAEPEDAESKRTYVARGPEPTPVVYGEKPS 136
Qy 121 KGDPTNTEIRQSDVEGDRDHRPQEKKAAGLGEITLMTQTLNTLSTPEKLAALCKKY 180
Db 137 KGDPTNTEIRQSDVEGDRDHRPQEKKAAGLGEITLMTQTLNTLSTPEKLAALCKKY 196
Qy 181 AELLIEHRNSQOMKLLQKKQSQLVQKDHLRGHSKAVLARSKLSLCRELQHRNSLK 240
Db 197 AELLIEHRNSQOMKLLQKKQSQLVQKDHLRGHSKAVLARSKLSLCRELQHRNSLK 256
Qy 241 BEGVQARAEEREEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300
Db 257 BEGVQARAEEREEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 316
Qy 301 ELRBEHDKVFKHDLQOOLVDAKLQQAQEMLKEAERHOREKDFLLKBAVESQRCMLM 360
Db 317 ELRBEHDKVFKHDLQOOLVDAKLQQAQEMLKEAERHOREKDFLLKBAVESQRCMLM 376

Qy 361 KQETHLKQALALYTEKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKLEKETTMTSRW 420
Db 377 KQETHLKQALALYTEKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKLEKETTMTSRW 436
Qy 421 ESSNKALLEWAEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQSSL 480
Db 437 ESSNKALLEWAEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQSSL 496
Qy 481 TDSGPERRPEGPAQAPSSPRVTAPCYPGAPSTEASGQTGPQPTTSARA 530
Db 497 TDSGPERRPEGPAQAPSSPRVTAPCYPGAPSTEASGQTGPQPTTSARA 546

RESULT 9
US-10-023-529-44
; Sequence 44, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-529-44

Query Match 99.7%; Score 2693; DB 13; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.2e-155;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 76
Qy 61 RQLEDILSTYCVDNNGGPGEDGAGGPAEPEDAESKRTYVARGPEPTPVVYGEKPS 120
Db 77 RQLEDILSTYCVDNNGGPGEDGAGGPAEPEDAESKRTYVARGPEPTPVVYGEKPS 136
Qy 121 KGDPTNTEIRQSDVEGDRDHRPQEKKAAGLGEITLMTQTLNTLSTPEKLAALCKKY 180
Db 137 KGDPTNTEIRQSDVEGDRDHRPQEKKAAGLGEITLMTQTLNTLSTPEKLAALCKKY 196
Qy 181 AELLIEHRNSQOMKLLQKKQSQLVQKDHLRGHSKAVLARSKLSLCRELQHRNSLK 240
Db 197 AELLIEHRNSQOMKLLQKKQSQLVQKDHLRGHSKAVLARSKLSLCRELQHRNSLK 256
Qy 241 BEGVQARAEEREEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300
Db 257 BEGVQARAEEREEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 316
Qy 301 ELRBEHDKVFKHDLQOOLVDAKLQQAQEMLKEAERHOREKDFLLKBAVESQRCMLM 360

Db 317 ELREHIDKVFVKHKLQQLVDKALQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM 376
QY 361 KOQETHLKOQLALYTEKPEPONTLSKSSSEVFTTFKQEMERMTKKIKLEKETTTMYRSRW 420
Db 377 KOQETHLKOQLALYTEKPEPONTLSKSSSEVFTTFKQEMERMTKKIKLEKETTTMYRSRW 436
QY 421 ESSNKALLEMAEETKVRDKLEGLQVKIQRLEKLCRALQTERNDLNKRVDLSAGGOSL 480
Db 437 ESSNKALLEMAEETKVRDKLEGLQVKIQRLEKLCRALQTERNDLNKRVDLSAGGOSL 496
QY 481 TDSGPERPPEGGAQAPSSPRVTEAPCYGAPSTEASGOTGPQEPSARA 530
Db 497 TDSGPERPPEGGAQAPSSPRVTEAPCYGAPSTEASGOTGPQEPSARA 546

RESULT 10
US-10-023-523-44
; Sequence 44, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR FILING DATE: 2000-03-02
; PRIOR FILING DATE: 1997-11-26
; PRIOR FILING DATE: 1996-11-27
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-523-44

Query Match 99.7%; Score 2693; DB 13; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.2e-155;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKPEGAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKPEGAQARTAQSGALRDVSEELS 76
QY 61 RQLEDILSTYCVDNQGGPDEGAQEPAPPEPAEAKSRVTVARNGEPPEPTPVVYGEKPS 120
Db 77 RQLEDILSTYCVDNQGGPDEGAQEPAPPEPAEAKSRVTVARNGEPPEPTPVVYGEKPS 136
QY 121 KGPNTETIRQSDVEGDRHRRPQEKKAQKGLGKEITLLMQTLNTLSTPEEKLAALCKKY 180
Db 137 KGPNTETIRQSDVEGDRHRRPQEKKAQKGLGKEITLLMQTLNTLSTPEEKLAALCKKY 196
QY 181 ABLEEHNSQKMLQKQKQSVOLVQKHDLRGEHSKAVLARSKLSLCRELQRHNSLK 240
Db 197 ABLEEHNSQKMLQKQKQSVOLVQKHDLRGEHSKAVLARSKLSLCRELQRHNSLK 256
QY 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNSKLRQENMELAEKLLIEQY 300
Db 257 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNSKLRQENMELAEKLLIEQY 316
QY 301 ELREHIDKVFVKHKLQQLVDKALQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM 360

Db 317 ELREHIDKVFVKHKLQQLVDKALQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM 376
QY 361 KOQETHLKOQLALYTEKPEPONTLSKSSSEVFTTFKQEMERMTKKIKLEKETTTMYRSRW 420
Db 377 KOQETHLKOQLALYTEKPEPONTLSKSSSEVFTTFKQEMERMTKKIKLEKETTTMYRSRW 436
QY 421 ESSNKALLEMAEETKVRDKLEGLQVKIQRLEKLCRALQTERNDLNKRVDLSAGGOSL 480
Db 437 ESSNKALLEMAEETKVRDKLEGLQVKIQRLEKLCRALQTERNDLNKRVDLSAGGOSL 496
QY 481 TDSGPERPPEGGAQAPSSPRVTEAPCYGAPSTEASGOTGPQEPSARA 530
Db 497 TDSGPERPPEGGAQAPSSPRVTEAPCYGAPSTEASGOTGPQEPSARA 546

RESULT 11
US-10-616-187-44
; Sequence 44, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-44

Query Match 99.7%; Score 2693; DB 15; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.2e-155;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKPEGAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKPEGAQARTAQSGALRDVSEELS 76
QY 61 RQLEDILSTYCVDNQGGPDEGAQEPAPPEPAEAKSRVTVARNGEPPEPTPVVYGEKPS 120
Db 77 RQLEDILSTYCVDNQGGPDEGAQEPAPPEPAEAKSRVTVARNGEPPEPTPVVYGEKPS 136
QY 121 KGPNTETIRQSDVEGDRHRRPQEKKAQKGLGKEITLLMQTLNTLSTPEEKLAALCKKY 180
Db 137 KGPNTETIRQSDVEGDRHRRPQEKKAQKGLGKEITLLMQTLNTLSTPEEKLAALCKKY 196
QY 181 ABLEEHNSQKMLQKQKQSVOLVQKHDLRGEHSKAVLARSKLSLCRELQRHNSLK 240
Db 197 ABLEEHNSQKMLQKQKQSVOLVQKHDLRGEHSKAVLARSKLSLCRELQRHNSLK 256
QY 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNSKLRQENMELAEKLLIEQY 300
Db 257 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNSKLRQENMELAEKLLIEQY 316

QY	1	KSSPQGPAGP	GAQERPSQAAPVAEAGPGSSQAPRKPRGQAQRTAQSGALRDVSEELUS	60
Db	17	KSSPQGPAGA	GAQGRPCRPAPAREAG-ASSQAPGRPEGAQTAQPGALCDVSEELUS	75
QY	61	ROLEDILSTY	CVDDNKGPGDGAQGEPAEPEDEAKSRTYVARNGPEP-TPVVYGEKBP	119
Db	76	ROLEDILSTY	CVDDNKGAPGEDGVQGEPEPEDEAKSRAYVARNGPEPQTPVNGEKET	135
QY	120	SKXGDPNTBE	IROSDVEGDRHRRPOEKKAKGLGKBITLLMQTLNTLSTPEEKLAALCKK	179
Db	136	SKAEPTGEBE	IRTSDEVGDRHRRPOEKKAKGLGKBITLLMQTLNTLSTPEEKLAALCKK	195
QY	180	YAELEEHNSO	KOMKLLQKQSOQIWOEKDHLRGHSAKVLARSKLESCLRELQNRSL	239
Db	196	YAELEEHNSO	KOMKLLQKQSOQIWOEKDHLRGHSAKILLARSKLESCLRELQNRSL	255
QY	240	KEEGVQRAEE	BEKEKEVTSHPQVTLNDIQLOEQHQRVNRNSKLRQENMELAEKLLIQ	299
Db	256	KEEGVQRAEE	BEKEKEVTSHPQVTLNDIQLOEQHQRVNRNSKLRQENMELAEKLLIQ	315
QY	300	YELREEHIDK	VPKHKDLOOQLVDAKLOAQEKVLKEAEERHQREXDQLLKEAVESQRMCEL	359
Db	316	YELREEHIDK	VPKHKDLOOQLVDAKLOAQEKVLKEAEERHQREXDQLLKEAVESQRMCEL	375
QY	360	MQQETHLKQO	ALYTEKPEBPONTLSKSSSEVFTTFKQEMEKMTKKIKLEKETTWYRSR	419
Db	376	MQQETHLKQO	ALYTEKEEFQNTLSKSSSEVFTTFKQEMEKMTKKIKLEKETTWYRSR	435
QY	420	WTSSNKALLE	MAEETKTVRDKLEGLQVKIQRLEKLCRALQTERNDLNKRVODLSAGQGS	479
Db	436	WTSSNKALLE	MAEETKLRDKLEGLQVKIQRLEKLCRALQTERNDLNKRVODLSAGQGS	495
QY	480	LTDGSGPERP	-----EGPQAQAPSSPRVTEAPCPVGPAPSTASGOTGQBPPTSA	528
Db	496	VDSGSGPERP	PEATTSKEQGVGPGQAVPNPRAITDASCAGASTASGOTGQBPPTSA	555
QY	529	RA	530	
Db	556	TA	557	

```

RESULT 14
US-10-671-242-5
; Sequence 5, Application US/10671242
; Publication No. US2004004049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; CURRENT FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus

```

```

US-10-671-242-5
Query Match          91.2%; Score 2465.5; DB 12; Length 557;
Best Local Similarity 90.8%; Pred. No. 8.3e-142;
Matches 492; Conservative 12; Mismatches 25; Indels 13; Gaps 3;

Qy 1 KSPGPGPEAGPEAGQERRPQAAPAVAAEAGPGSSQAPKPEGAQARTAQSGALRDVSELS 60
    ||||| ||||| ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 17 KSPGPGPEAGPEAGQERRPQAPAPRABG-ASSQAPRPEGAQAKTAQPGALCDVSELS 75
    ||||| ||||| ||||| : ||||| : ||||| : ||||| : ||||| : |||||

Qy 61 RQLEDILSTCYVDNNQGGPCGDEGAQGEPAEPEDAERKSTRTVARNGPEP-TPVYVGEKEP 119
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 76 RQLEDILSTCYVDNNQAGPCGDVGQGEPEPEPEDAEKSRAYVARNGPEPEPTPVVNGEKE 135
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 120 SKGDPNTEIRQSDVEGDRDHRPQRKKKAGLGGKEITLLMOTLNTLSTPEEKLAALCKK 179
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 136 SKAAPTETETIRSDVEGDRDHRPQPKKAGLGGKEITLLMOTLNTLSTPEEKLAALCKK 195
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 180 YALLIEEHNSQOMKLLQKQSQLVQEKDHLRGEHSKAVLARSKLESCLRELQRHNSL 239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 196 YALLIEEHNSQOMKLLQKQSQLVQEKDHLRGEHSKAILARSKLESCLRELQRHNSL 255
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 240 KEKGVORAREEHKREKVTSHFQVTLNDTQLQNEQHNRNSKLQENMELAEELKKLIHQ 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256 KEKGVORAREEHKREKVTSHFQVTLNDTQLQNEQHNRNSKLQENMELAEELKKLIHQ 315
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 300 YELREHIDKVPKHKDLQQLVDKLLQQAQEMLKAEERHOREKDFLLKEAVESQRMCEL 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 316 YELREHIDKVPKHKDLQQLVDKLLQQAQEMLKAEERHOREKDFLLKEAVESQRMCEL 375
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 360 MKQOETHLQKQALYKTEKTEBPONTLSKSEVTPTPKQEMEKWTKIKKLEKTTMYRSR 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 376 MKQOETHLQKQALYKTEKTEBPONTLSKSEVTPTPKQEMEKWTKIKKLEKTTMYRSR 435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 420 WESSNALLMEABEKTVRDKEGLGVQKIQRLKLCRALQTERNDLMKRVQDLSAGQSGS 479
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 436 WESSNALLMEABEKTVRDKEGLGVQKIQRLKLCRALQTERNDLMKRVQDLSAGQSGP 495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 480 LITDGGERRP-----EGFGAQFSSPRVTEAPCYGPAPSTASQGVQPEPTSA 528
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 496 VSDSGPERRPEPATTSKSGQVEGPGQVPSPRATDASCCAGAPSTASQGVQPEPTTA 555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 529 RA 530
    |
Db 556 TA 557

RESULT 15
US-10-623-529-5
; Sequence 5, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING AND TREATING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53

```

; SOFTWARE: FastSeq for Windows Version 4.0			; PRIOR APPLICATION NUMBER: US 60/031,930		
; SEQ ID NO 5			; PRIOR FILING DATE: 1996-11-27		
; LENGTH: 557			; PRIOR APPLICATION NUMBER: US 60/048,547		
; TYPE: PRT			; PRIOR FILING DATE: 1997-06-03		
; ORGANISM: Oryctolagus cuniculus			; NUMBER OF SEQ ID NOS: 53		
US-10-023-529-5			; SOFTWARE: FastSeq for Windows Version 4.0		
Query Match			; SEQ ID NO 5		
Best Local Similarity 91.2%; Score 2465.5; DB 13; Length 557;			; LENGTH: 557		
Matches 492; Conservative 12; Mismatches 25; Indels 13; Gaps 3;			; TYPE: PRT		
			; ORGANISM: Oryctolagus cuniculus		
			US-10-023-523-5		
QY	1	KSSPGQPEAGPEGAQERPSQAAPAEVAGPGSSQAPRKPEGAQARTAQSGALRDVSEELS 60	QY	1	KSSPGQPEAGPEGAQERPSQAAPAEVAGPGSSQAPRKPEGAQARTAQSGALRDVSEELS 60
Db	17	KSSPGQPEAGPEGAQERPSQAAPAEVAGPGSSQAPRKPEGAQARTAQSGALRDVSEELS 75	Db	17	KSSPGQPEAGPEGAQERPSQAAPAEVAGPGSSQAPRKPEGAQARTAQSGALRDVSEELS 75
QY	61	ROLEIDILSTYCVNNQGGPDGAGQEPAPEDAKSRRTYVARNGEP-TPVYGEKEP 119	QY	61	ROLEIDILSTYCVNNQGGPDGAGQEPAPEDAKSRRTYVARNGEP-TPVYGEKEP 119
Db	76	ROLEIDILSTYCVNNQGGPDGAGQEPAPEDAKSRRTYVARNGEP-TPVYGEKEP 135	Db	76	ROLEIDILSTYCVNNQGGPDGAGQEPAPEDAKSRRTYVARNGEP-TPVYGEKEP 135
QY	120	SKGDPNTEIRQSDVGDHRRPOEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKK 179	QY	120	SKGDPNTEIRQSDVGDHRRPOEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKK 179
Db	136	SKAEPGTEIRTSDEVDGRDHRPPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKK 195	Db	136	SKAEPGTEIRTSDEVDGRDHRPPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKK 195
QY	180	YAEILLEHRNSQKQMLLQKQSQOLVQEKDHLRGEHSAVLARSKLSLCRELQRNRS 239	QY	180	YAEILLEHRNSQKQMLLQKQSQOLVQEKDHLRGEHSAVLARSKLSLCRELQRNRS 239
Db	196	YAEILLEHRNSQKQMLLQKQSQOLVQEKDHLRGEHSAVLARSKLSLCRELQRNRS 255	Db	196	YAEILLEHRNSQKQMLLQKQSQOLVQEKDHLRGEHSAVLARSKLSLCRELQRNRS 255
QY	240	KEEGVORAREEERKEVTSHPQVTLNDIQLQMEQHNEHNSKLRQENMELAEKLLIEQ 299	QY	240	KEEGVORAREEERKEVTSHPQVTLNDIQLQMEQHNEHNSKLRQENMELAEKLLIEQ 299
Db	256	KEEGVORAREEERKEVTSHPQVTLNDIQLQMEQHNEHNSKLRQENMELAEKLLIEQ 315	Db	256	KEEGVORAREEERKEVTSHPQVTLNDIQLQMEQHNEHNSKLRQENMELAEKLLIEQ 315
QY	300	YELREEHIDKVFVKHDLQQLVDKLLQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 359	QY	300	YELREEHIDKVFVKHDLQQLVDKLLQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 359
Db	316	YELREEHIDKVFVKHDLQQLVDKLLQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 375	Db	316	YELREEHIDKVFVKHDLQQLVDKLLQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 375
QY	360	MKQETHLKOOLALYTEKEFEFONTLSSSVFTTFKQEMKMTKKIKLEKETTYRSR 419	QY	360	MKQETHLKOOLALYTEKEFEFONTLSSSVFTTFKQEMKMTKKIKLEKETTYRSR 419
Db	376	MKQETHLKOOLALYTEKEFEFONTLSSSVFTTFKQEMKMTKKIKLEKETTYRSR 435	Db	376	MKQETHLKOOLALYTEKEFEFONTLSSSVFTTFKQEMKMTKKIKLEKETTYRSR 435
QY	420	WESSNKALLEWAEKTLKDEGLQVKIQRLEKLCRALQTERNDLNKRVQDL-SAGQGS 479	QY	420	WESSNKALLEWAEKTLKDEGLQVKIQRLEKLCRALQTERNDLNKRVQDL-SAGQGS 479
Db	436	WESSNKALLEWAEKTLKDEGLQVKIQRLEKLCRALQTERNDLNKRVQDL-SAGQGS 495	Db	436	WESSNKALLEWAEKTLKDEGLQVKIQRLEKLCRALQTERNDLNKRVQDL-SAGQGS 495
QY	480	LTDSPERRP-----EGPGAQAPSSPRVTRAPCYPGAPSTASQGTGPQETSA 528	QY	480	LTDSPERRP-----EGPGAQAPSSPRVTRAPCYPGAPSTASQGTGPQETSA 528
Db	496	VSDSGPERRPPEATTSTKEQGVGPGAGVPSNPSPRATDASCCAGAPSTASQGTGPQETTA 555	Db	496	VSDSGPERRPPEATTSTKEQGVGPGAGVPSNPSPRATDASCCAGAPSTASQGTGPQETTA 555
QY	529	RA 530	QY	529	RA 530
Db	556	TA 557	Db	556	TA 557
RESULT 16			RESULT 17		
US-10-023-523-5			US-10-616-187-5		
; Sequence 5, Application US/10023523			; Sequence 5, Application US/10616187		
; Publication No. US20020152485A1			; Publication No. US20040013668A1		
; GENERAL INFORMATION:			; GENERAL INFORMATION:		
; APPLICANT: Lees, Ann M.			; APPLICANT: Lees, Ann M.		
; APPLICANT: Lees, Robert S.			; APPLICANT: Lees, Robert S.		
; APPLICANT: Law, Simon W.			; APPLICANT: Law, Simon W.		
; APPLICANT: Arjona, Anibal A.			; APPLICANT: Arjona, Anibal A.		
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING			; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING		
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING			; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING		
; FILE REFERENCE: 10797-004001			; FILE REFERENCE: 10797-004001		
; CURRENT APPLICATION NUMBER: US/10/023,523			; CURRENT APPLICATION NUMBER: US/10/616,187		
; CURRENT FILING DATE: 2001-12-17			; CURRENT FILING DATE: 2003-07-09		
; PRIOR APPLICATION NUMBER: US/09/616,289			; PRIOR APPLICATION NUMBER: US/09/616,187		
; PRIOR FILING DATE: 2000-07-14			; PRIOR FILING DATE: 2003-07-09		
; PRIOR APPLICATION NUMBER: US 09/517,849			; PRIOR APPLICATION NUMBER: US 08/979,608		
; PRIOR FILING DATE: 2000-03-02			; PRIOR FILING DATE: 1997-11-26		
; PRIOR APPLICATION NUMBER: US 08/979,608			; PRIOR APPLICATION NUMBER: US/09/616,289		
; PRIOR FILING DATE: 1997-11-26					

```

, PRIOR FILING DATE: 2000-07-14
, PRIOR APPLICATION NUMBER: US 09/517,849
, PRIOR FILING DATE: 2000-03-02
, PRIOR APPLICATION NUMBER: US 08/979,608
, PRIOR FILING DATE: 1997-11-26
, PRIOR APPLICATION NUMBER: US 60/031,930
, PRIOR FILING DATE: 1996-11-27
, PRIOR APPLICATION NUMBER: US 60/048,547
, PRIOR FILING DATE: 1997-06-03
, NUMBER OF SEQ ID NOS: 53
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 5
, LENGTH: 557
, TYPE: PRT
, ORGANISM: Oryctolagus cuniculus
US-10-636-187-5

```

```

RESULT 18
US-10-276-774-2134
, Sequence 2134, Application US/10276774
, Publication NO. US20040053245A1
, GENERAL INFORMATION:
, APPLICANT: Hysec, Inc.
, APPLICANT: Tang, Y. Tom et al
, TITLE OF INVENTION: NO. US20040053245
, FILE REFERENCE: 21272-030
, CURRENT APPLICATION NUMBER: US/10-276
, CURRENT FILING DATE: 2003-11-18

```

```

; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2134
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc:feature
; LOCATION: (1)..(510)
; OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-2134

```

```

, RESULT 19
, US-10-425-114-55188
, Sequence 55188, Application US/10425114
, Publication No. US20040034888A1
, GENERAL INFORMATION:
, APPLICANT: Liu, Jingdong
, APPLICANT: Zhou, Yihua
, APPLICANT: Kovalic, David K.
, APPLICANT: Screen, Steven E.
, APPLICANT: Tabaska, Jack E
, APPLICANT: Cao, Yongwei
, TITLE OF INVENTION: Nucleic Acid Molec
, TITLE OF INVENTION: Plants and Uses T
, FILE REFERENCE: 38-21(53313)B
, CURRENT APPLICATION NUMBER: US/10/425.
, CURRENT FILING DATE: 2003-04-28
, NUMBER OF SEQ ID NOS: 73128
, SEQ ID NO 55188
, LENGTH: 437
, TYPE: PRT

```


187 HRNSQOMKLOKQSQVOEKDHLRGEHSKAVLARSKLESCLRELQRHNSRLKEGVOR 246
31 HSKLEKDKELF-----QLAQYHKVVAERDQATAVKORLESCLREFORQNKYLKEEQV 85
247 AREBEKREKVTSHQVTLNDIQLQM-EQNNRNSKLRQNMELARLKKLIEQYELREB 305
86 STEGQNNRTLSKFKHAIKGVSAKLEBQERISQL-BENNTLRSLKDLADQYNIQO 144
306 HIDKVFHKDLOOQVNDAKLQQAQEMLKEAEERHOREKOPFLKEAVESQRMCELMKQOET 365
145 KYAHOLKEREKLELADLRLOQHQE--KAAQHTQWQ-----LYAEQVSQLM-TTEK 193
366 HLKQOALALYKTEPEEFQNTLSKSEVFTTQKQEMKMTKKIKLEKETMYRSRWSSNK 425
194 NLRLQASDCGERFQHDALSKSENFETTKQEMKMSVINKLKENEFKGCNSDI 253
426 ALLEMAEKTVRDKEGLEQVQKRLKCLRALQTERNDLNKRVQDLSAGGQSLTDSGP 485
254 AIVKLEERELTKQIEKLNQREKLESCLRTQAEK----- 291
486 ERPEGFGAQPSPRYTE 504
292 ----QGFSASIPDAPSSQE 306

RESULT 22
US-10-369-493-5130
; Sequence 5130, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)3
; CURRENT APPLICATION NUMBER: US/10369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5130
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5130

Query Match 13.9%; Score 375; DB 15; Length 346;
Best Local Similarity 30.8%; Pred. No. 5.4e-15;
Matches 105; Conservative 69; Mismatches 123; Indels 44; Gaps 9;

QY 150 KGLGKEITLMQ-----TLNTL-STPE-EKLAALCHQYAELEEHNSOKMKLQKQ 201
DB 3 KNFGKLSIFNEDGAALLKLEGVDAEKVKQKLIKGLAE--SEKQNAELKIKVLDYK 60
QY 202 SQVQEKDHLRGEHSKAVL-----ARSKLESCLRELQRHNSRLKEGVORAREBEKRE 256
DB 61 VVKVQDTEKLERNOILRTEBKSLRELRCGLQKXAHQTRERACAKMKLEVERGL 120
QY 257 VTSHQVTLNDIQLQMEQNNRNSKLRQNMELARLKKLIEQYELREB-----LREE 305
DB 121 AVEQLKVTLDIKETMAEGRSKSDLAEDNKKLSEKFGHGYEERKMKVIDQIQKCK 180
QY 306 HIDKVFHKDLOOQVNDAKLQQAQEMLKEAEERHOREKOPFLKEAVESQRMCELMKQOET 365
DB 181 WYEEYGTQKLETKLTKLESASIQVKKG-----GMEKDELAKIMLEETARVQKALTEK 236
QY 366 HLKQOALALYKTEPEEFQNTLSKSEVFTTQKQEMKMTKKIKLEKETMYRSRWSSNK 425
DB 237 ALREQVEYSAKYSELTSCLSKSNEAFDKFKDEISRVNKKCMQVKEGLSKKKSDBANK 296

426 ALL-----EMAEKTVRDKEGLEQVQKRLKCLRALQTERNDLNKRVQDLSAGGQSLTDSGP 459
297 KVLVTMTNOEYAEKIATSDK-----KIQMLENLCLRALR 330

RESULT 23
US-10-425-114-45419
; Sequence 45419, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 45419
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3088-030-A1_FLI.pep
US-10-425-114-45419

Query Match 11.3%; Score 304.5; DB 12; Length 206;
Best Local Similarity 34.5%; Pred. No. 5.5e-11;
Matches 78; Conservative 48; Mismatches 77; Indels 23; Gaps 6;

QY 273 EQNNRNSKLRQNMELARLKKLIEQYELREBHDKVKFKHDLQOOLVDAKLQQAQEML 332
DB 4 EQRVETCIAQLENNN-LRSKLDIADQYNIQOYAHQKERNMELELADLRLOQHQE-- 60
QY 333 KEAEERHOREKOPFLKEAVESQRMCELMKQOETHLQOALALYKTEPEEFQNTLSKSEV 392
DB 61 KAAHEHTQWQ-----LYAEQVSQLM-TTEKRLQLASGGERFQHDALSKSEV 111
QY 393 TTFQKQEMKMTKKIKLEKETMYRSRWSSNKALLEMAEKTVRDKEGLEQVQKRL 452
DB 112 ETVKQEMKMTLVIKNLKKNDFLKGKENSIALVKLIEERELATKQIEKLNQKQETLE 171
QY 453 KLCRALQTERNDLNKRVQDLSAGGQSLTDSGPERPBGPGAQPS 498
DB 172 SLCSLQAEKQ-----GRSASIPDA-PSSQEDMPATSQES 206

RESULT 24
US-10-369-493-1016
; Sequence 1016, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1016
; LENGTH: 886
; TYPE: PRT

```
; ORGANISM: Archaeoglobus fulgidus
US-10-359-493-1016

Query Match      9.6%; Score 259; DB 15; Length 886;
Best Local Similarity 24.8%; Pred. No. 1.9e-07;
Matches 113; Conservative 86; Mismatches 111; Indels 146; Gaps 22;

QY 116 EKPSKGDPTNTEETIQSDVCDRHRPPOEKKAGLCKEITLLMOTLNTLSTPEKLA 175
DB 171 EKEELK-----EFLSQEQQ-----KQKEKKK-----ELRISSEIKSISREKLS- 214

QY 176 LCKYAELEHRNSQOMKLLQKQSQOLVQBDHLRGHSAVLARSKLESLECRELQRH 235
DB 215 -----EBVRNLSRLKELEHKSRL-----ESLAKQESSVLQEVVRGLEEKLEBLE-- 259

QY 236 NRSLKESGVQARSEEEKKEV-----TSHFQVTLND----- 267
DB 260 -KQKEV-VERIEBLEKKAKEVKELPKAERYSTILEKLLSEINQALRDVKEKRGDLTRE 317

QY 268 --IQLOMEQHNRNSKLQENMELARLKKLIEQY----- 303
DB 318 AGIQALCKAEDNSKL-BEITTKRIEELERLEFERFKSHRLLETLPKMDRMQGIKAKLE 376

QY 304 EEHI--DKVPFKHKLQOQVDAKLOQAQEMLKEABERHOR-----EKDFL-----LKEAVE 352
DB 377 EKNUTPKVEQWYDL-----LSKAKEBEKEITEKLKLIACKSSLKTRGALCKAVE 428

QY 353 ----SORMCELMQO--ETHLKQOLALYTRKPFQNTLSKSSSEVFTTFKQEMKMTK 406
DB 429 ELKSAERTCPVCGRELDSEHRKNIMAEVTRMKIAESLAKADEIEKKLXERLEKVE--- 485

QY 407 KLEKETMTWRSYH-----ESNKALLEMAEKVTVRDKELSL--QVKI-- 448
DB 486 KALEKQETVLKQWMDLKALENELSHDPAEKLSAESEYRKVKYER-LDGLRGQOKILL 544

QY 449 -----ORLEKLCRALQTERNDLNKRVOD 471
DB 545 SSASRIKELSSLRIEEALKNWESERGELFKIRE 580

RESULT 25
US-10-072-012-838
; Sequence 838, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zethusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Beha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCES: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 838
; LENGTH: 860
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Myosin tail
US-10-072-012-838

Query Match      9.5%; Score 257.5; DB 12; Length 860;
Best Local Similarity 22.7%; Pred. No. 2.3e-07;
Matches 143; Conservative 106; Mismatches 189; Indels 193; Gaps 25;

QY 12 EGAQRPSQAAFAVAEAGPGSSQAPRKPEGAQARTAQSGALRDVSEELSRQLEDILSTYC 71
DB 77 EELSERLEAGGATAAQ-----IELNKKRAELAKLAKLDLEENLQHEALAT-L 125

QY 72 VDNQGGPEDGAQGEPAPEPAEAKSRVTVARNGPBPPTPVVYGEKPSKGDPTTEE-IR 130
DB 126 RKKHQDAINELSEQIEQLQKQAK-----AEKBSQLQAEVDDLLA 166

QY 131 QSDVEGDRDHRPPOEKKAGLCKEIT-----LLMOTLNTLSTPEKLA-- 175
DB 167 QLDSEI---TKAKLNAEKKAKQLESQSLSELOVKLDELQNLDTQSKSELOSENSDLITQ 223

QY 176 -----LCKYAEI---LEHRNS-----QKMKLLQKQSQOLVQBDHLRGHSA 218
DB 224 LEEBAQVSNLSKLKLSQLESQLEAKRSLEESRERANLQALRQLEHDLDSLREGLEEE 283

QY 219 VLARSKLESLECEIQRHNRSLKE-----EGVQARSEEEKKEV--- 258
DB 284 SEAKAELE---RQLSKANABIQWRSKFSESGALRAESELKLLKLNKIKISLEBAABAA 340

QY 259 -----SHFQVTLNDIQLOMEQHNRNSKL--RQENNE--LAERLKKLIE----- 298
DB 341 NAKCDSLKTKSRLQSELEDLQILERANAAASELEKQKQKPFKILAEWRKRVDELQABL 400

QY 299 ---OYELR-----EEHDKV---FKHKDLOQQLVDAKLOQAQ---EMLK 333
DB 401 DTAQREARNLSTELPRLKNELEELKQVQVEALRRKNLQDEIHDLTQDQGEGRNVHELE 460

QY 334 EAEERHOREKDFLLKEAVESQRMCEL-----MKQETHLKQOLALYTEKPEEFQ 382
DB 461 KARRLEAKKDELQALAEARALLESKVLRAQVELSQIRSEIERLLA---EKESEPE 517

QY 383 NT-----LSKSSSEVFTTFKQEMKMTK-----IK 407
DB 518 NTRKNHORAISLOATLEATKGAESRLKKQLEGDINELEIALDHANKANAEAKQNVK 577
QY 408 KLEKETMTWRSWSSNKA-----LLEMAEEK-TVRDKELEGLOVKIORLEKLCRALQ 460
DB 578 KYQQQVKELQTVBEHQAREDAEQLAVAEERRATALEAELEELRSALSAEQAEARKQAE 637
```

QY	461	ERNDLNKRVQDLSAGGQGSILTSGPERRREG	491
Db	638	ELAEASERVNLTQNSSLIAQ---	KKLEGG 665
RESULT 26			
US-10-037-417-59			
; Sequence 59, Application US/10037417			
; Publication No. US20040052806A1			
; GENERAL INFORMATION:			
; APPLICANT: Kekuda, Ramesh			
; APPLICANT: Alsobrook II, John P			
; APPLICANT: Tchernev, Velizar T			
; APPLICANT: Liu, Xiaohong			
; APPLICANT: Spytek, Kimberly A			
; APPLICANT: Patturajan, Meera			
; APPLICANT: Grosse, William M			
; APPLICANT: Lepley, Denise X			
; APPLICANT: Burgess, Catherine E			
; APPLICANT: Vernet, Corine A.M.			
; APPLICANT: Li, Li			
; APPLICANT: Gorman, Linda			
; APPLICANT: Edinger, Shlomit R			
; APPLICANT: Sciore, Paul			
; APPLICANT: Ellerman, Karen			
; APPLICANT: Malyankar, Uriel M			
; APPLICANT: Rotherberg, Mark			
; APPLICANT: Stone, David J			
; APPLICANT: Boldog, Ferenc L			
; APPLICANT: Guo, Xiaojia			
; APPLICANT: Shenoy, Suresh G			
; APPLICANT: Anderson, David W			
; APPLICANT: Padigaru, Muralidhara			
; APPLICANT: Taupier Jr, Raymond J			
; APPLICANT: Miller, Charles E			
; APPLICANT: Eisen, Andrew J			
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same			
; FILE REFERENCE: 21402-235			
; CURRENT APPLICATION NUMBER: US/10/037,417			
; CURRENT FILING DATE: 2002-09-20			
; PRIOR FILING DATE: 2002-09-20			
; PRIOR FILING DATE: 2001-01-05			
; PRIOR FILING DATE: 2001-01-05			
; PRIOR FILING DATE: 2001-01-08			
; PRIOR FILING DATE: 2001-01-08			
; PRIOR FILING DATE: 2001-02-28			
; PRIOR FILING DATE: 2001-02-28			
; PRIOR FILING DATE: 2001-03-02			
; PRIOR FILING DATE: 2001-03-02			
; PRIOR FILING DATE: 2001-05-15			
; PRIOR FILING DATE: 2001-05-15			
; PRIOR FILING DATE: 2001-07-05			
; PRIOR FILING DATE: 2001-07-05			
; PRIOR FILING DATE: 2001-07-12			
; PRIOR FILING DATE: 2001-07-12			
; PRIOR FILING DATE: 2001-09-10			
; PRIOR FILING DATE: 2001-09-10			
; PRIOR FILING DATE: 2001-09-12			
; NUMBER OF SEQ ID NOS: 227			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 59			
; LENGTH: 860			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence: Mysosin Tail			
; OTHER INFORMATION: Consensus Sequence			
US-10-037-417-59			
Query Match			
Best Local Similarity 22.7%; DB 12; Length 860;			
Matches 143; Conservative 106; Mismatches 189; Indels 193; Gaps 25;			

QY	12	EGAOERPSQAAPVAEAGPGSSQAPRKPEGAQARTAQSGALRDVSEBELSRQLEILSTVC	71
Db	77	BELSERLEBEAGGATAAQ-----	JELNKRREAEALAKRLKDLBEANLQHEEALAT-L 125
QY	72	VDNQGGGCHDGAQCHPEAPBPDAEKSRVTVANGSEPTPVVYGSKEPSKCDPNTTEE-IR	130
Db	126	RKKGHDAINELSEQTEQLQKQKAK-----	AEKESQLQAEVDDLLA 166
QY	131	QSDVEYCDRHRPQEKKAAGLGKEIT-----	ILLQTLNTLSTPPEKLA- 175
Db	167	OLDSI---TKAKLNAEKKAQLESQLESLQVVKLDELQRLNDLTSQKSLQSENSDLTRQ	223
QY	176	-----LCKYABL---LEHRNS-----	OKQKLLQKQSOVLQVQEKHGRHSKA 218
Db	224	LEEAQVSNLSKLSQLESQLEEAQRSLEBSREANLQAQRLQLEHDLDSLRQLEEE	283
QY	219	VLARSKLSLCLRELQHRNSLKE-----	EGVQARBEERKEKRV- 258
Db	284	SEAKAELE---ROLSKANAEIQWTSKPESEGALEAELELKKLNQKISLEBAEAA	340
QY	259	-----SHFQVTLNDIQLOVQENRNSKL--ROENVE--LAERLKKLIE-----	298
Db	341	NAKCDSEKTKSRQSELEDLQIELERANAAASELEKQKQKDFKILAEWKRVDBLOAEL	400
QY	299	---QYSLR-----SEHIDKV---FHKDLQOQLVDAKLQQAQ-----	EMLK 333
Db	401	DTAQREARNLSTELFRLKVELSELQVQVEALPRENKNLQDEIHDLTDQLGEGGRNVHELE	460
QY	334	EAEERHQRKDFLLKEAVESQRMCEL-----	MKQOETHLKKQALYTKPEEPQ 382
Db	461	KARRLEAKDELQAALBEARAALEBSKVLRQAQVELSQIRSEIERLIA---EKESPE	517
QY	383	NT-----LSKSSSVFTTFKQEMEKMTKK-----	IK 407
Db	518	NTRKNHQAIBESLQATLEATKGAASRLKKKLGSDINELEIALDHANKANAQKQNVK	577
QY	408	KLEKETTYRSWESSNKA-----LLEAEK-TVRDKLEGLQVKVQIRLKLCLRALQT	460
Db	578	KYQQQVGEIQTQVEBEQARERDAREQLAVABERRATALEAELELSALSALQAEARAKQAST	637
QY	461	ERNDLNKRVQDLSAGGQGSILTSGPERRREG	491
Db	638	ELAEASERVNLTQNSSLIAQ---	KKLEGG 665
RESULT 27			
US-10-080-334-166			
; Sequence 166, Application US/10080334			
; Publication No. US20040002584A1			
; GENERAL INFORMATION:			
; APPLICANT: Pena, Carol E. A.			
; APPLICANT: Shimkets, Richard A			
; APPLICANT: Li, Li			
; APPLICANT: Shenoy, Suresh G			
; APPLICANT: Kekuda, Ramesh			
; APPLICANT: Spytek, Kimberly A			
; APPLICANT: Vernet, Corine A. M.			
; APPLICANT: Malyankar, Uriel M			
; APPLICANT: Guo, Xiaojia			
; APPLICANT: Gusev, Vladimir Y			
; APPLICANT: Casman, Stacie J			
; APPLICANT: Boldog, Ferenc L			
; APPLICANT: Furtak, Katarzyna			
; APPLICANT: Tchernev, Velizar T			
; APPLICANT: Patturajan, Meera			
; APPLICANT: Gangolli, Esha A			
; APPLICANT: Padigaru, Muralidhara			
; APPLICANT: Liu, Xiaohong			
; APPLICANT: Baumgartner, Jason C.			
; APPLICANT: Gerlach, Valerie			
; APPLICANT: Spaderna, Steven X			
; APPLICANT: Zerhusen, Bryan D			

QY 299 ---QVELR-----BEHIDKV-----FKHDLQOQLVDAKLQAAQ-----EMLK 333
Db 401 DTAQREARNLSTELFLRANELEELKQVFEALRRKNKQDEIHDLTDQGEGRNVHELE 460
QY 334 EAEERHOREKDFLKEAVESQSMCEL-----MKQOETHLKKOOLALYTEKPEFQ 382
Db 461 KARRLEAEKDELQALAEAEAALEESKVLRAQVELSQIRSEIERELA---EKESFPE 517
QY 383 NT-----LSKSEVFTTFKQEMKMKK-----IK 407
Db 518 NTRKHQRAIESLOATLEAETKGAENSLRKKKLEGDINELEIALDHANKANAQAQKVK 577
QY 408 KLEKETTMYRSWESSNKA-----LLEWABEK-TVRDKLEGLQVQIKLEKLCALQT 460
Db 578 KYQQVKELQVQVEEQEQRADARQLAVARRATALEAELEELRSALAEQAERARKQABT 637
QY 461 ERNDLKNRVQDLSAGGQSLTDSGERRPBG 491
Db 638 ELAEASERVNLTACNSSLIAQ---KRLKLG 665

RESULT 28
US-10-146-473-41
; Sequence 41, Application US/10146473
; Publication NO. US2003010888A1
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Gout, Ivan
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Gure, Ali
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Old, Lloyd
; TITLE OF INVENTION: Breast Cancer Antigens
; FILE REFERENCE: L00461/70130(JRV)
; CURRENT APPLICATION NUMBER: US/10/146,473
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/291,150
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 41
; LENGTH: 2871
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-146-473-41

Query Match 9.5%; Score 256; DB 14; Length 2871;
Best Local Similarity 23.2%; Pred. No. 1.2e-06;
Matches 133; Conservative 107; Mismatches 195; Indels 138; Gaps 25;

QY 26 EAEFGSSQAPRKP-----EGQAQRTAQS-----GALRDVSELSRQLEDIL 67
Db 1432 QKATSGSEVSQKQLEVLRLQVTOVRTTESTYKOSLDDAKTIQDKKEIER-LKQLI 1490
QY 68 STYCVNNQGGGEGDGAQGEPAEPDAKSRT-YVARNGEPSTPVV-----YGEKPSKG 122
Db 1491 DKETNDR-----KCLEENARLQVQYDLQKANSATETINKLVQEQELTRL 1538
QY 123 DPWTEIRQSDVGRDRHRPQKKAQGLGKEITLMTLNTLSTPEEKLAALCKYAE 182
Db 1539 RIDYERSQRTYQDDITRFQNSLKELQKQKQ-KVEEELNKLKRTASDSCRKKLE 1596
QY 183 LLEHRNSQKMKL-----LQKQSQ--LVQEKD-----HLR-GEHSKAVLAR 222
Db 1597 ELEGRSLKEQAIKITNLTLQQLSQASIVKRSDDLQQRDVLGDHLEKQRTQELRR 1656
QY 223 --SKLESICHELQHRNLSKE-----EGVQRAEEBEKKEVTSHFQVTLNDIQLQWEQH 275
Db 1657 LSSEVALRQLIQBQSVQKQHLRNEHFQKATEDKS-----SUNESKIEERL 1706
QY 276 NERNKLRQNMELAEERLKKL-LEQVELR-----EZHDK-----VFKHDLQ 318
Db 1707 QSLTENITKHEHMLEELNRLLEYDDLRGRSEADSKNATILELSRQISNNRTLEL 1766

QY 12 EGAQRPQAAQAVAEAGFGSQAPRKPQGAQRTAQAQSGALRDVSELSRQLEDILSTYC 71
Db 77 EELSRLBEAGGATAAQ-----TELNKGAEALAKLRKDLLEANLQHEALAT-L 125
QY 72 VDNQGGPGEGDGAQGEPAEPDAKSRTYVARNGEPEPTPVYGEKPSKGDPTHE-IR 130
Db 126 RKHQDALNELSEIEQLEQKQKAK-----AEKESQLOAEVDDLLA 166
QY 131 QSDVGRDRHRPQKKAQGLGKEIT-----LLMOTLNTLSTPEEKLA 175
Db 167 QLDSI---TKAKLNAAEKKAQLESQSLQELQDLQRLNDLTSQKRSQSENSDLTRQ 223
QY 176 -----LCKYAEGL---LLEHRNS-----QKMKLQKQSQLOVCKDHLRGEHKA 218
Db 224 LEEAQAQVNSLKLQKLESQLESQLEAKRSLRESREANLQQLRQLEHDLQSLAEQLE 283
QY 219 VLARSLKSLCHELQHRNLSKE-----EGVQRAEEBEKKEVTSHFQVTLNDIQLQWEQH 258
Db 284 SEAKAELE---ROLSKANAEIQWRSKFSEGAALRAEELKLNQKINSELSEAEAEA 340
QY 259 -----SHFQVTLNDIQLQWEQHNRSLK---ROBANE---LAERLKKLIE----- 298
Db 341 NAKCDSLENTKSRQSELEDIQLIELERANAAASELKKQKQNFDKILAEWKKVDELQAE 400

Query Match 9.5%; Score 257.5; DB 15; Length 850;
Best Local Similarity 22.7%; Pred. No. 2.3e-07;
Matches 143; Conservative 106; Mismatches 189; Indels 193; Gaps 25;

QY 12 EGAQRPQAAQAVAEAGFGSQAPRKPQGAQRTAQAQSGALRDVSELSRQLEDILSTYC 71
Db 77 EELSRLBEAGGATAAQ-----TELNKGAEALAKLRKDLLEANLQHEALAT-L 125
QY 72 VDNQGGPGEGDGAQGEPAEPDAKSRTYVARNGEPEPTPVYGEKPSKGDPTHE-IR 130
Db 126 RKHQDALNELSEIEQLEQKQKAK-----AEKESQLOAEVDDLLA 166
QY 131 QSDVGRDRHRPQKKAQGLGKEIT-----LLMOTLNTLSTPEEKLA 175
Db 167 QLDSI---TKAKLNAAEKKAQLESQSLQELQDLQRLNDLTSQKRSQSENSDLTRQ 223
QY 176 -----LCKYAEGL---LLEHRNS-----QKMKLQKQSQLOVCKDHLRGEHKA 218
Db 224 LEEAQAQVNSLKLQKLESQLESQLEAKRSLRESREANLQQLRQLEHDLQSLAEQLE 283
QY 219 VLARSLKSLCHELQHRNLSKE-----EGVQRAEEBEKKEVTSHFQVTLNDIQLQWEQH 258
Db 284 SEAKAELE---ROLSKANAEIQWRSKFSEGAALRAEELKLNQKINSELSEAEAEA 340
QY 259 -----SHFQVTLNDIQLQWEQHNRSLK---ROBANE---LAERLKKLIE----- 298
Db 341 NAKCDSLENTKSRQSELEDIQLIELERANAAASELKKQKQNFDKILAEWKKVDELQAE 400

QY	319	QYVDAKLOQAQEMLKABEBBHQEKOPLLKBAVESQRMCLMKQOQBTHLKQQLALVTEK	377
DB	1767	QGLINDLOREBNROBIEKFQKALAEASNRIQESKNQCTQVVOERBSLLVKKVLEQCK	1826
QY	378	-----FEFQNTLSKSSEVTTFKQEMEKTKKIKKLEKETTMYASRMSSNKALIEKA	431
DB	1827	ARLQRLDELINRAKSTLEARTVQKRLCEKQIQONLQNKWTQYSPKEAIRK-IESE	1884
QY	432	EKTVRDK-----ELEGLOVKIORLEKLCR-----ALQTERDNLKRVQDLSA	474
DB	1885	REKSERKNSLRSEIERLQAEIKRIBERCARKLEDSTRETQSQLEIERSYQREIDKL--	1942
QY	475	GGQSLTDSQFERRPPGGAQAPSSPVITEAPC	507
DB	1943	-----RQPPYC-----SHREQTQEC	1957

```

RESULT 29
US-10-032-585-7646
; Sequence 7646, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Rosmer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NC 7646
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: {1881}..(1881)
; OTHER INFORMATION: X-any amino acid
US-10-032-585-7646

```

[illegible]

Db 1127 SMI AKLSAKIEBHK-----TDEIETKTKHTDLOEEHAKQSFESERNDIK 1175

Qy 429 EMAEENTVEDKELEGLOVKIORLEKLCLRALQTERNDLNKRVO 470
 :||| : ||| : ||| : ||| : ||| :
Db 1176 SNLDE---ANKELSDNRKSLNLEK-----EKTELANKLKX 1207

RESULT 30

US-10-369-493-21643

? Sequence 21643, Application US/10369493

? Publication No. US20030233675A1

? GENERAL INFORMATION:

? APPLICANT: Cao, Yongwei

? APPLICANT: Hinkle, Gregory J.

? APPLICANT: Slater, Steven C.

? APPLICANT: Goldman, Barry S.

? APPLICANT: Chen, Xianfeng

? TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

? FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

? FILE REFERENCE: 38-10(52052)B

? CURRENT APPLICATION NUMBER: US/10/369,493

? CURRENT FILING DATE: 2003-02-28

? PRIOR APPLICATION NUMBER: US 60/360,039

? PRIOR FILING DATE: 2003-02-21

? NUMBER OF SEQ ID NOS: 47374

? SEQ ID NO 21643

? LENGTH: 880

? TYPE: PRT

? ORGANISM: Pyrococcus abyssi

US-10-369-493-21643

Query Match	9.3%; Score 252.5; DB 15; Length 880;
Best Local Similarity	23.2%; Pred. No. 4.7e-07;
Matches	123; Conservative 98; Mismatches 183; Indels 127; Gaps 22;
QY	12 EGAQRPSSQAAPAVAEAGPGSSQ-----APRKPEGQAQTAQSGALRDVSEELSRULE 64
DB	259 XGLEEKIVQIERSIEKKAKISLSEIBIVDIPLOQEKKEVRLKG-PRDEVESKLRLLE 317
QY	65 DILSTYCDVNNQGGPEGGAQEPAPEDAERSRTTYVARNGSEPPTFVVYGEKPEPSKDP 124
DB	318 KELSKN-----ESELKAISBEVIKEGB-----KKKE 342
QY	125 NTBEIRQS-DEVGDR-DHRRP--OEKQKAGLGKIHITLLMQTLINTLSTPE--EKLAALCK 178
DB	343 RABEIRKLSIEBKRLLEELKPYVELEDAKOVQOIBRLKARLKSLSPGEVIEKLESLEK 402
QY	179 KYAEELLEHNSSOKMKLQKXSQLOVQEKDHLRGHSHKAVLARKLSLSC-----REL-Q 233
DB	403 ERTEI-----BEALKEITTRIGQHEQKN-----ERKAIIBELKAKGKCPVCGRELTE 451
QY	234 RHRSRLK-----EEGVORAREEBEKKREKVTSHFQVTLNDIQLQ-----MEQHNERN 280
DB	452 EHKKELMERYTLLEIKLIEBELKRTTEERKLVNRLKLEIKLREFSVMRDIAEQIKEL 511
QY	281 KLRQENMELAE-----RUKGLTEQVELREHHDKVPK 312
DB	512 KLAGFNLELEQKEREPEGLNBEFNKLGELGLERDLKRIKALBGRKLIIEKVRKAK 571
QY	313 H-KDLOOLVDAKLQAQEN---LKEAERHOREXDFLLKEAVESQRMCELMKQOETELK 368
DB	572 ELENLHQLBELGFESVEELNLSIQLEEFPHDKYVEAKKSES--ELRELKMKLEKTELD 630
QY	369 QQLALVTEKPEBPQNTLSKSSVEFTTF--KQENEXMTKKIKLEKETTMYVRSRWBSNKA- 426
DB	631 QAFEMLADVNEIEEKEAKLKOLESKFNEEYEEKREKRLVLEREVSSUTARLEBLKKS 690
QY	427 -----ILEMAEBKTVRDKELGLQVQIKRLKLCALQTERNDLNKRVQD 471
DB	691 EQIKATLTKLKBKBEKREK-----AKLEIKKALKALSKE-----DLRKKID 733

RESULT 31

US-09-291-417-23
; Publication No. US020030050230A1
; Sequence 23, Application US/09291417A
; Publication No. US020030050230A1
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 240/300
; CURRENT APPLICATION NUMBER: US/09/291.417A
; EARLIER FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: US 60/081,784
; EARLIER FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Mammalian (Human) SULU3
US-09-291-417-23

Query Match 9.2%; Score 248.5; DB 10; Length 786;
Best Local Similarity 21.6%; Pred. No. 7.2e-07;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

Qy 23 PAVAEAE-----GPGSSQA-PRKPEGAQAARTAGSALRDVSEELSRQLE 64
Db 109 PAVAEAEEDHGVGRGTGVNSVGSNQSIIPMSISASSOSSVNSLPDVSDDKS-ELD 167
Qy 65 DILSTYCYDNNQG---GPGSDG--AOGEPAPEDAEKRTTYVARNGEPEFTPVVYGEKE 118
Db 168 MMEGDITVMSNSSVIHLKPEENYREGDP-----RT---RASDPQSPQVSRHKS 215
Qy 119 PSKGDPTTEIRQSDVEGDRHRRPOEKKAKGLGKEIT---LLMQTLNTLSTPEKLA 174
Db 216 HYRNRHFATIRTSLV---TQMQEHEQDSLELRQSGYKMRMRQHQKQMLTENK 271
Qy 175 ALCKKYA---ELLEHRN--SQQMKLLQKQSQLVQE-----KDLRGHSHK 217
Db 272 AEMDEHRLDKDLETQNNFAEMKLIKHOAAMEKAKVMSNEEKFOQHIQAQOKK 331
Qy 218 AVLARKSLKSLCRLQHRNRSKKEGVOR---AREBEE---KKEVTSHPQV--TINDI 268
Db 332 EL--NSFLESQKREYKLRKEQLKEELNENQSTPKKEQEWLSKQENIQHFOAEENLL 389
Qy 269 QLOME-----QHN-----ERNSKLRQENMELAEKLLKLIQYELRE 304
Db 330 PRQRYLELCERFPRMMLGRHNLQDLVREELNKRQTKDLEHMLLRQESMQELEP 449
Qy 305 EHDKVPKHDQLOOLVDKLAQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELMKQOE 364
Db 450 RHNTTIQK---MRCELI--RLQHOTELTNQLEYNKRERELRKHVMEVVRQOPKSLKSKE 504
Qy 365 THLKQOL---ALYTEKEFEFQNTLSKSESVFTTFKQEMEKMTKKIKKLEKETTYMSRW 420
Db 505 LQIKKQFQDTCKIQTRQYKALRNHLE---TTPKSEHKAVLRK--EBOTRKLAILA 557
Qy 421 ESSNKALLEMAEBKTVR-----DKELE 442
Db 558 EYDHSINEMLSQALRLDERQAEBCQVLKMQLOQHELELLNAYQSKIKMQAEAQHRELR 617
Qy 443 GLOVKI-----ORLEKLCRALQTERND-----LNKRVODLSAGGGS-----L 480
Db 618 ELEQVRSRLRALLEOKIEEMALQWTERIRSLERQAREIEAFDSMRGFSNMVL 677
Qy 481 TDSGPERPEPGGAQAPSSPRVTEAPCYPGAP--STEASGOTGP 522
Db 678 SNLSPE-----AFSHSYFGASGWSHNPTGGPG 705

US-09-291-417-31
; Publication No. US20030050230A1
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 240/300
; CURRENT APPLICATION NUMBER: US/09/291.417A
; EARLIER FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: US 60/081,784
; EARLIER FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Mammalian (murine/human) SULU3
US-09-291-417-31

Query Match 9.2%; Score 248.5; DB 10; Length 1001;
Best Local Similarity 21.6%; Pred. No. 9.6e-07;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

Qy 23 PAVAEAE-----GPGSSQA-PRKPEGAQAARTAGSALRDVSEELSRQLE 64
Db 324 PAVAEAEEDHGVGRGTGVNSVGSNQSIIPMSISASSOSSVNSLPDVSDDKS-ELD 382
Qy 65 DILSTYCYDNNQG---GPGSDG--AOGEPAPEDAEKRTTYVARNGEPEFTPVVYGEKE 118
Db 383 MMEGDITVMSNSSVIHLKPEENYREGDP-----RT---RASDPQSPQVSRHKS 430
Qy 119 PSKGDPTTEIRQSDVEGDRHRRPOEKKAKGLGKEIT---LLMQTLNTLSTPEKLA 174
Db 431 HYRNRHFATIRTSLV---TQMQEHEQDSLELRQSGYKMRMRQHQKQMLTENK 486
Qy 175 ALCKKYA---ELLEHRN--SQQMKLLQKQSQLVQE-----KDLRGHSHK 217
Db 487 AEMDEHRLDKDLETQNNFAEMKLIKHOAAMEKAKVMSNEEKFOQHIQAQOKK 546
Qy 218 AVLARKSLKSLCRLQHRNRSKKEGVOR---AREBEE---KKEVTSHPQV--TINDI 268
Db 547 EL--NSFLESQKREYKLRKEQLKEELNENQSTPKKEQEWLSKQENIQHFOAEENLL 604
Qy 269 QLOME-----QHN-----ERNSKLRQENMELAEKLLKLIQYELRE 304
Db 605 PRQRYLELCERFPRMMLGRHNLQDLVREELNKRQTKDLEHMLLRQESMQELEP 664
Qy 305 EHDKVPKHDQLOOLVDKLAQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELMKQOE 364
Db 665 RHNTTIQK---MRCELI--RLQHOTELTNQLEYNKRERELRKHVMEVVRQOPKSLKSKE 719
Qy 365 THLKQOL---ALYTEKEFEFQNTLSKSESVFTTFKQEMEKMTKKIKKLEKETTYMSRW 420
Db 720 LQIKKQFQDTCKIQTRQYKALRNHLE---TTPKSEHKAVLRK--EBOTRKLAILA 772
Qy 421 ESSNKALLEMAEBKTVR-----DKELE 442
Db 773 EYDHSINEMLSQALRLDERQAEBCQVLKMQLOQHELELLNAYQSKIKMQAEAQHRELR 832
Qy 443 GLOVKI-----ORLEKLCRALQTERND-----LNKRVODLSAGGGS-----L 480
Db 833 ELEQVRSRLRALLEOKIEEMALQWTERIRSLERQAREIEAFDSMRGFSNMVL 892
Qy 481 TDSGPERPEPGGAQAPSSPRVTEAPCYPGAP--STEASGOTGP 522
Db 893 SNLSPE-----AFSHSYFGASGWSHNPTGGPG 920

RESULT 33

US-10-177-293-254

; Sequence 254, Application US/10177293

; Publication No. US20030124128A1

; GENERAL INFORMATION:

RESULT 32

US-09-291-417-31

; Sequence 31, Application US/09291417A

APPLICANT: Lillie, James
APPLICANT: Glat, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Gannavarpu, Manjula
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Mertens, Maureen
APPLICANT: Myer, Vic
APPLICANT: Wang, Youzhen
APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: East Jr., Robert C.
APPLICANT: Hoshobasyi, Gabriel N.
APPLICANT: Pusztai, Lajos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegul
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038
CURRENT APPLICATION NUMBER: US/10/177,293
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 254
LENGTH: 1005
TYPE: PRT
ORGANISM: Homo sapiens
US-10-177-293-254

Query Match 9.2%; Score 248.5; DB 14; Length 1005;
Best Local Similarity 21.6%; Pred. No. 9.7e-07;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;
Qy 23 PAVEAE-----GPGSQA-PRKPEGAQARTAGALRDVSELSQLE 64
Db 328 PAVEAE-----GPGSQA-PRKPEGAQARTAGALRDVSELSQLE 64
Qy 65 DILSTYCVNNQ-----GPGEDG--AQGEPAEPEDAESKRTTYVARNGEPPTPVYGEKE 118
Db 387 MMEGDHTVMSNSSVHLKPEENYREBGP-----RT--RASDPQSPQVSRHKS 434
Qy 119 PSKGDPTTEIROSDVGRDHRPQEKKAGLKEIT-----LLMQLTLSTPBEKLA 174
Db 435 HYRNREHPATIRTSLV-----TROMQEHEQDSLEQMSGYKMRQHQKQATLENKLA 490
Qy 175 ALCKKYA-----ELLEHRN--SOKMKLQKQSQVLQV-----KDLRGHSHK 217
Db 491 AEMDEHRLDKDLETQNNFAAEKLIKQQAAMEKAKVMSNEEKKQHQIAQOKK 550
Qy 218 AVLARKLSLCRLQHRNRLKKEGVOR-----AREEEE---KRKVTSHFQV--TLNDI 268
Db 551 EL--NSFLSKQREYKLRKEQLKEELNENQSTPKKEQEWLSKQENIQHFQAEFEANLL 608
Qy 269 QLOQE-----QHN-----ERNSKLRQENMELAEKLLKLIQYELRE 304
Db 609 RRQRYLELCRRFRKRMMLGRHNLQDLVREELNKRQTKQLEHMLLRHSHESQLEF 668
Qy 305 EHDVKFKKDKLQQLVDKILQQAQEMKAEERHREKDFLLKEAVESQRMCELMKQOE 364
Db 669 RHLNTTQK--MRCELI--RLQHTQELTNQLEYNKRRERELRKHVMEVQPPKSLKSKE 723
Qy 365 THLKQQL-----ALYTEKFEFQNTLSKSEVFTTFKQEMERMTTKIKKLEKETTMYRSRW 420

Db 669 RHLNTTQK--MRCELI--RLQHTQELTNQLEYNKRRERELRKHVMEVQPPKSLKSKE 723
Qy 365 THLKQQL-----ALYTEKFEFQNTLSKSEVFTTFKQEMERMTTKIKKLEKETTMYRSRW 420
Db 724 LQIKKQFQDTCKIQTRQYKALRNHLE-----TTPKSEHKAVLKRLK--EEQTRKLAILA 776
Qy 421 ESSNKALLEKABEKTVR-----DYELE 442
Db 777 EYDHSINEKLSQALRLDEAQBACQVLKMQLOQLLELLNAYQSKIKMQAQAQHDRELR 836
Qy 443 GLOVKI-----QRLEKLCALQTERND-----LNKRVQDLSAGQGS-----L 480
Db 837 ELQORVSLRALLEQKLEEMALQNERTERINSLLERQAREIEAFDSESMRLGPSNMWL 896
Qy 481 TDSGPRRPEGCAQAPSSPRVTEAPCYGAP--STEASGQTOP 522
Db 897 SNLSPE-----AFSHSYFGASGWSHNP7GGPG 924
RESULT 34
US-10-317-835-18
; Sequence 18, Application US/10317835
; Publication No. US20030157531A1
; GENERAL INFORMATION:
; APPLICANT: Exelixis, Inc.
; TITLE OF INVENTION: TAOJIKS AS MODIFIERS OF THE BETA-CATENIN PATHWAY AND METHODS OF
; FILE REFERENCE: EX02-141C
; CURRENT APPLICATION NUMBER: US/10/317,835
; CURRENT FILING DATE: 2002-12-12
; PRIOR APPLICATION NUMBER: 60/340,312
; PRIOR FILING DATE: 2001-12-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-317-835-18

Query Match 9.2%; Score 248.5; DB 14; Length 1005;
Best Local Similarity 21.6%; Pred. No. 9.7e-07;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

Qy 23 PAVEAE-----GPGSQA-PRKPEGAQARTAGALRDVSELSQLE 64
Db 328 PAVEAE-----GPGSQA-PRKPEGAQARTAGALRDVSELSQLE 64
Qy 65 DILSTYCVNNQ-----GPGEDG--AQGEPAEPEDAESKRTTYVARNGEPPTPVYGEKE 118
Db 387 MMEGDHTVMSNSSVHLKPEENYREBGP-----RT--RASDPQSPQVSRHKS 434
Qy 119 PSKGDPTTEIROSDVGRDHRPQEKKAGLKEIT-----LLMQLTLSTPBEKLA 174
Db 435 HYRNREHPATIRTSLV-----TROMQEHEQDSLEQMSGYKMRQHQKQATLENKLA 490
Qy 175 ALCKKYA-----ELLEHRN--SOKMKLQKQSQVLQV-----KDLRGHSHK 217
Db 491 AEMDEHRLDKDLETQNNFAAEKLIKQQAAMEKAKVMSNEEKKQHQIAQOKK 550
Qy 218 AVLARKLSLCRLQHRNRLKKEGVOR-----AREEEE---KRKVTSHFQV--TLNDI 268
Db 551 EL--NSFLSKQREYKLRKEQLKEELNENQSTPKKEQEWLSKQENIQHFQAEFEANLL 608
Qy 269 QLOQE-----QHN-----ERNSKLRQENMELAEKLLKLIQYELRE 304
Db 609 RRQRYLELCRRFRKRMMLGRHNLQDLVREELNKRQTKQLEHMLLRHSHESQLEF 668
Qy 305 EHDVKFKKDKLQQLVDKILQQAQEMKAEERHREKDFLLKEAVESQRMCELMKQOE 364
Db 669 RHLNTTQK--MRCELI--RLQHTQELTNQLEYNKRRERELRKHVMEVQPPKSLKSKE 723
Qy 365 THLKQQL-----ALYTEKFEFQNTLSKSEVFTTFKQEMERMTTKIKKLEKETTMYRSRW 420

Db 724 LOIKQFQDTCIKIOTROYKALRNHLE-----TTPKSEHKAVALKRLK--BEQTRKLAIA 776
Qy 421 ESSNKALLEMAEETVR-----DKELE 442
Db 777 EYDHSINEMLSQALRLDEAQAECQVLYKQQLQOELELNAVQSKIKMQAQAQHDREL 836
Qy 443 GLQVKI-----QRLKLCRALQTERND-----LNKRVODLSAGQGS-----L 480
Db 837 ELEQVSLRALLQKIEEMALQNERTERIRSLRQARETEAFDSMRLGFSNMVL 896
Qy 481 TDSGPERRPEGQAQAFSSPRVTEAPCPGAP--STEASGQTGP 522
Db 897 SNLSPE-----APSHSYPGASGWSHNPCTGPGP 924

RESULT 35
US-10-049-3413
; Sequence 3413, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 3413
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: MISC_FEATURE
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (3)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (75)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-049-3413

Query Match 9.1%; Score 247; DB 15; Length 85;
Best Local Similarity 64.5%; Pred. No. 5.8e-08;
Matches 49; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

Qy 347 LKAVESORMCMLAKQOETHLKKQALALYKTEFEQNTLSKSEVFTTFQEMKMTKKI 406
Db 2 LKAVESRMKVEQMKQEVQLKKQLSLYMDKPEEFQTTMAKSNELVTPPQEMKMTKKI 61
Qy 407 KLEKETMYRSRWES 422
Db 62 KLEKSTIWRKXET 77

RESULT 36
US-10-445-735-2
; Sequence 2, Application US/10445735
; Publication No. US20040009175A1
; GENERAL INFORMATION:
; APPLICANT: Cobb, Melanie
; TITLE OF INVENTION: TAO PROTEIN KINASE POLYPEPTIDES AND METHODS OF USE THEREOF

; FILE REFERENCE: 10624-026-999
; CURRENT APPLICATION NUMBER: US/10/445,735
; CURRENT FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: US/09/686,346
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 09/060,410
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; US-10-445-735-2

Query Match 9.1%; Score 247; DB 15; Length 1001;
Best Local Similarity 21.9%; Pred. No. 1.2e-06;
Matches 142; Conservative 118; Mismatches 212; Indels 176; Gaps 30;

Qy 23 PAVEAE-----GPGSSQA-PRKPEGQAQTAQSGALDVSEELSRQLE 64
Db 324 PAVEAQEEBEEHGGGRTGTVNSVGSNQSIISMISASSQSSSVNSLPDASDDKS-ELD 382
Qy 65 DILSTYCVDNQO-----GFGDGAQCEPAEPDEDAEKSRITYVARNGEPEPTPVVYGEKPS 120
Db 383 MMEGDHTVMSNSSVIHLKEEENYQ--EKGDP-----TRASAPQSPQVSRHSHY 432
Qy 121 KQDPNTEIRQSDVGDHRRPQEKKAKGLGKEIT-----LLMQTLNTLSPEKLAAL 176
Db 433 RNREHFATIRTSALV-----TROMQEHQDSELRQMSGYKRMRRQHQKQMTLENKLKAB 488
Qy 177 CKKYA-----ELLEERN--SOKOMLKQKQQLVQE-----KDLHRGEHSAV 219
Db 489 MDEHRLDLXOLETORNNFAEMEKLIKHOASMEKAVMANEKKFOHQAQOQKEL 548
Qy 220 LARSKLSLCRELQHRNRLKEGVQR-----AREEBE---KKVTSHPQV--TLNDIOL 270
Db 549 --NSFLESQREYKLRKEQLKEELNENOSTPKKEQEWLSKQENIKOHFAEEANLLR 606
Qy 271 QME-----OHN-----BRNSKLQENMELAEKRLKLIQYELREBH 306
Db 607 QRYLELCRFRFRMLGRHMLQDLVREELNKQQTQDLEHMLLRQESMQELEPRH 666
Qy 307 IDKVFHKDLQOQLVDKILQQAQEMLEKAEERHOREKDFLLKAEVSEORMCMLKQOETH 366
Db 667 LNTIQK---MRCELI--RLQHQTELNTQLEYNKREERELRRKHVMVEVRQOPKSLKSLQ 721
Qy 367 LKQOL-----ALYKTEFEFQNTLSKSEVFTTFQEMKMTKKIKLEKETMYRSRWES 422
Db 722 IKQFQDTCIKIOTROYKALRNHLE-----TTPKSEHKAVALKRLK--BEQTRKLAIA 774
Qy 423 SNKALLEMAEETVR-----DKELEGL 444
Db 775 YDHSINEMLSQALRLDEAQAECQVLYKQQLQOELELNAVQSKIKMQAQAQHDREL 834
Qy 445 QVKI-----QRLKLCRALQTERND-----LNKRVODLSAGQGS-----LTD 482
Db 835 BORVSLRALLQKIEEMALQNERTERIRSLRQARETEAFDSMRLGFSNMVL 894
Qy 483 SGFERRPEG-PCAQAPS--SPRVTEAP--CYPGAPSTEASG---QTCGP 523
Db 895 LSPFAFSHSYFGASSSHNPCTGSGPHNGHPMGCTPQAWHPMQGPGP 942

RESULT 37
US-09-925-299-988
; Sequence 988, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299

```

; CURRENT FILING DATE: 2001-03-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 988
; LENGTH: 909
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (32)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (41)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (47)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (48)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (52)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (58)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (62)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (125)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (632)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (851)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-925-299-988
; Query Match
; Best Local Similarity 22.4%; Score 244.5; DB 9; Length 909;
; Matches 151; Conservative 109; Mismatches 237; Indels 177; Gaps 32;
; QY 1 KSSPGQFEAG--PEGQAEPSQAAPAVE--AEGPGSSQAPRPEGAQTAQ-----48
; DB 132 KAEGAGNQKKAEGAGNQKKAEGAGNQKQKGEAGNQKKGTEGAQKKAERSPNQK 191
; QY 49 -----SGALRVSELSRQLEDILSTYCDVNNQGGFBGAGQEPAPEDAB-----K 96
; DB 192 GEGAPIQKKADSVANQGTKEGI-----TNQG-----KKAEGSPSEGGKAEKSPNQK 240
; QY 97 SRTYVARNPEPTPVYVYKEKSPKGDPTNTEBROSDEVDGRDRHPQEKK---KAKG--151
; DB 241 KADAAANQKKTBSAQVR-----NT-DVAQSPAPKQKQ--APAKKSGSKKKGEP 289
; QY 152 -----LGEKITLAKQTLNLTSTPE---EKLAALCKKYASLLER--HRNSQK--QMKL 196
; DB 290 GPPDADGPLYLPYKTLVSTGVSMVNEGEAQLIETLSKAGIITDTHKATQKGPVAI 349
; QY 197 LQKKQSQLVQKDLRGEHSKAVLAKSLKSLCRL--QHRNRLKEGVQR---ARESE 251
; DB 350 LKR---QLEEKELLATQEDAAVAKSKLRELKEMAAEKAKAAAGAEKVKQLVAREGE 406
; QY 252 -----EKKEVTSHPQVTLNDIQLQMEQ--HNERNSKLQENNELAERLKLI 297
; DB 407 ITAVQRMQASREHVKV--QQLQKIRTLQELNGPNTQLARLQKENSILDALNQAT 465
; QY 298 EQYE-----LREHIDKVPK-----HFDLQQLVDAKLQQAQEMLEAEER 338
; DB 466 SQVESKQNAELAKLQKQ--LSKSVKSELVKSAYRQDEQKALEKAAAFKQVLOQAS 524
; QY 339 HOREKDFLLKEAVESQRMCELMKQOETHLKQOLALYTKFEFFQNTLS-----KSSEVP 392
; DB 525 HRESEALQKRLDEVR--ELCHTQSSH--ASLRADAERKAEQEQQQAELHSHKLSQSEAE 580
; QY 393 TTFK-----QEMEKNTKKIKKLEKETTMYRSRWESSNKALLEMAEKTVR 437
; DB 581 VRSKCELSGLHGLOEBAEANSQUTERIRSEALLEAGADQADQVQASQXEAQOQTR 640
; QY 438 DKELE---GLQVKIQRLKLCRALQTERNDLN-----KRVQDLSAGGQ-----SLTDS 483
; DB 641 LKELESQVSGLEKEALELREAVQKQKNDLRKNWKAMEALATAEQACKELHSLTQA 700
; QY 484 GPERP-----EGPGAQA-----PSSPRVTEAPCYGAPST 514
; DB 701 KESEKQICLIRAEQMTMEALLLPELSVLAQONYTEWLODLKEKGTLLKHPAPAPBSS 760
; QY 515 BASGQTGPQEPSTA 528
; DB 761 DLASKLREAEQTQS 774
; RESULT 38
; US-09-925-299-988
; Sequence 988, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 988
; LENGTH: 909
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (32)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (41)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (47)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (48)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (52)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (58)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (62)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (125)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (632)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
```

LOCATION: (851)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-988

Query Match 9.0%; Score 244.5; DB 10; Length 909;
Best Local Similarity 22.4%; Pred. No. 1.5e-06;
Matches 151; Conservative 109; Mismatches 237; Indels 177; Gaps 32;

QY 1 KSPGQPEAG--PEGAPRPSQAPAVE--AEGPGSSQAPRPEGGAQARTAQ-----48
DB 132 KXAEQAQNGKKAEGHQAQNGKKAEGQAQNGKQKTEGQAQKKAERSPNQKK 191
QY 49 -----SGALRDVSEBLSRQLEDILSTYCVNNQGGPBGDGAQGPAPEDAR-----K 96
DB 192 GEGAPIQKKAADVANGTKVEGI-----TNQG-----KKAEGSPSEKKAEGSPNQG 240
QY 97 SRYTVARNGEPEPTPVVYGESEKSKGDPNTEBTRQSDVGDHRRPEQKK---KAKG--151
DB 241 KADAAANQKKAERSASVQGR-----NT-DVAQSPAPKQ--APAKKSKKKGEP 289
QY 152 -----LQKEITLLMOTLNTLSTPE---EKLAALCKKYAEELLE--HRNSQK--QMKL 196
DB 290 GPPDADGPLYLYPYKTLVSTVGSVMFNEGEAQLRILSEKAGIIOPTHKATQKGDPAI 349
QY 197 LQKQSQVLQVQKHLRGEHSHKAVLARSKLESICREL--QHNRSLKEEGVOR---AREEE 251
DB 350 LXR---QLESEKKLILATEQBDAAVAKSKURELNKEMAAEKAKAAAGAEKVKQLVAREQE 406
QY 252 -----EKKEVTSHPQVTLNDIQLQMEQ--HNRNSKLQENMELAEKLI 297
DB 407 ITAVQARMQASVREHYKEV--QOLQKIRTIQEQLENGPNQTLARLQENSILRDALNQAT 465
QY 298 EQVE-----LREHIDKVKF-----HKDLOQLVDKLOQAQEMKEAER 338
DB 466 SOVESKQNAELAKLRQ--LSKVSKELVKSEAVRQDEQQRKALEAKAAAFKQVLQAS 524
QY 339 HOREKDFLLKEAVESQRMCLMQQETHLQQAALVTEKPEPNTLS-----KSEVP 392
DB 525 HRSEELQKLEDEVR--ELCHTQSH--ASLRADAERKAEQEQQVABLSKLSSEAE 580
QY 393 TTFK-----QEMEMTKKIKLEKETTMYRSWESSNKVALLAEKTVR 437
DB 581 VRSKCEBLSGLHGOLEPARAENSQLTIRISIEALLEAGQARDQADQVQASXEAQOQTR 640
QY 438 DKELE-----GLOVKIQRLKCLALOTERNDLN---KRVODLSAGQG-----SLTDS 483
DB 641 LKELESQVGLKEKAEIAREAVQKVKYNDLREKNWKAMEALATAEQACKEXLHSLTQA 700
QY 484 GPERRP-----EGPGAQA-----PSSPRVTAPCPGAPST 514
DB 701 KESEKOLCLITRAQTEAMEALLALLPESLVLAQNVTEWLQDKKGPPTLKHPAPAPSPS 760
QY 515 EASQQTGPQETSA 528
DB 761 DLASKLREABETQS 774

RESULT 39

US-10-342-136-1
Sequence 1, Application US/10342136
Publication No. US20030199092A1
GENERAL INFORMATION:
APPLICANT: Meyer, David I.
TITLE OF INVENTION: METHOD FOR INCREASING MRNA HALF-LIFE IN EUKARYOTIC CELLS
FILE REFERENCE: 407T-301610US
CURRENT APPLICATION NUMBER: US/10/342,136
CURRENT FILING DATE: 2003-01-13
PRIOR APPLICATION NUMBER: 60/347,533
PRIOR FILING DATE: 2002-01-11
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1
LENGTH: 1534

TYPE: PRT
ORGANISM: Canis familiaris
US-10-342-136-1

Query Match 9.0%; Score 242; DB 14; Length 1534;
Best Local Similarity 20.4%; Pred. No. 4e-06;
Matches 153; Conservative 114; Mismatches 222; Indels 260; Gaps 30;

QY 1 KSPGQPEAGPE-----GAQERPSQAPAVEAGPGSSQAP 36
DB 710 KKEGSPNQGKVKVDASANQSKRAESAPITQGNADVMVQSEAPKQEPAPAKTKSGSKGGE 769
QY 37 RKP-----EQAQAR-----TAQSCALRDVSEB-----58
DB 770 GPPDSPLPYLYPYKTLVSTVGSVMFNEGEAQLRILSEKAGIIOPTHKATQKGDPAI 829
QY 59 LSRQLED---ILSTYCVNNQGGPBGDGAQGPAPEDAKSRT---YVARNGEPEPTPV 112
DB 830 LKRQLEKEKLLAT-----EQEDAAVAKSKLREVNKELAAKAKA 869
QY 113 VYGEKEPSK---GDPNTEETRSQSDVGDHRRPEQKKAK-----G 151
DB 870 AAGEAKVKQLVARQEIITAVQARIEASYREHVEVQOQKIRTLQOLENGPNQTLAR 929
QY 152 LQKEITLLMOTLN-----TLSTPEKLAALCKKYAEELLEHRNSQKMKLQK 199
DB 930 LQENSILRDALNQATSOVESKQNTLAKRLQELSKVSKELVEKSEAAARQEQKALBT 989
QY 200 KQSQVLQVQKHLRGEHSHKAVLA-RSKLESICREL---QHNRSLKEEGVORAREEEXK 255
DB 990 KTAALEKQVLQASHKESSEALQGLDEVSRELCRSQTSHASLRAD-AEKAQEQOQMA 1048
QY 256 EYTSHPQVTLNDIQLQMEQHNRNSKL---ROENNELAEKLLK---LIEQVLELRE---304
DB 1049 ELHSLKQSEAEVKSSELSLHQLKEARANSOLMERINSIEALLEAGQARDQADQ 1108
QY 305 -----EHIDKVKFKHDLQOOLVDKLOQAQEMKEAERHQ-----REKDFLLKEAVES-353
DB 1109 ASRAEHQARL---KXLESQV---WCLEKEATELKEAVEQKVKYNDLREKNWKAMEALASA 1163
QY 354 QMCEILMKQETHLQKQ-----LAL-----YTE-----376
DB 1164 ERACEKLSLTQAKESEKQLSLTEAQTEKALLALLPALSSAPQSYTEWLQELREKGP 1223
QY 377 -----KFEFQNTLSKSEVFTTFKQEMKMTKIKK-LEKETT 414
DB 1224 ELLKQRPADTDPSSDLASKLREABETQNNLQAECDQVRTILAEETGLMLDLQKSYEEBQ 1283
QY 415 MYRSWESSNKVALLAEKTVRDKLESLQVKIQ-----RLEKLCR 456
DB 1284 V---WKAKVSATEBELOKSRVTVKHLSDIVKELGELSSQVREHSHLEAELEKMA 1339
QY 457 ALQTERNDLNKRVQDLSAGQGSGLTDSGPERPEPGGAQAPSS-----PRVTEAPCY--508
DB 1340 AASACQSVAKV-----AGLRQLLESQSLDAAKSEAKQSQNELALVRQOLSEKSHVE 1395
QY 509 -----PGAPSTHAGQGTGPQETSA 530
DB 1396 DGDVAGSPAAPPAAE-----QDPVELKA 1417

RESULT 40

US-10-205-219-102
Sequence 102, Application US/10205219
Publication No. US20030138803A1
GENERAL INFORMATION:
APPLICANT: Warner-Lambert Company
APPLICANT: Lee, Kevin
APPLICANT: Dixon, Alistair
APPLICANT: Brooksbank, Robert
APPLICANT: Finnock, Robert
TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
FILE REFERENCE: WL-A-018200

```

; CURRENT APPLICATION NUMBER: US/10/205,219
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 102
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Myosin heavy chain
US-10-205-219-102

Query Match
Best Local Similarity 8.8%; Score 238; DB 14; Length 892;
Matches 113; Conservative 88; Mismatches 159; Indels 164; Gaps 18;

QY 33 SQAPKPGCAQRTAQSALDVSEHLSRQLEDILSTYCVNNQGGPGGAGGAPAEPE 92
DB 45 SNLQSKIEDQALGNQ---LQKIKELQARIEEL-----EEIEASRASRA 87
QY 93 DAEKSRITYVANGSGPEPTPVVYGEKPSKGDENTEEIRQS-DEVDGRDHRPQEKKAKG 151
DB 88 KAEKQSDLSR-----ELEEISLERLEAGGATSQIENWKKEA 126
QY 152 LQKEITLLMOTINTLSTPEKLAALCK-----VAELLEHRNSQKMKLLQKQSQLVQ 206
DB 127 ---EFQKRRDLEATLQHEATAATLAKKHADSVAEELGEQIDNLRVQKLEKEKSEMGM 183
QY 207 EKDHGRGSHKAVLARSK--LESLEELQ-----233
DB 184 EIDDLAS--NMEVSKSGNLEKMCRTLEDOVSELKTEBEQORLINELTAAQRGLQTES 241
QY 234 -RHNSLKE-----EGVQARAREEERKEVTSK-FQVTLNDIQLQW 272
DB 242 GEYSRQLEKDSVLSQSRGKQAFQTOQIEELKRLQLEBEVKAASALAHASSHDCDLR 301
QY 273 ECHNER-----NSKLR-----QENNELAEELKLEIQEVELBEHI 307
DB 302 EYVEEQEAKAELQRAMSKANSEVAQWTKYETDAIQSTEELEBAKKLAQRLQDAEHV 361
QY 308 KVPFKHDLQOOLVDKALQQAEMLKAEERHQBKDFLKEAVESQRMCKMQQETHL 367
DB 362 -----EAVNKAAS-----LEKTKORLQNEVEDLMIDVERTNAACAALDKQRNF 406
QY 368 KQALALYTEKPEEFQNTLSKSSEVFTTPFKQEMERXMTKKIKKLEKTTMYRSWESSNKAL 427
DB 407 DKLLAEWKQKYETH-----AELEASQKESRSISTELPKINAVERSLDQL 452
QY 428 LEMAEKTVRDKLEGL-QV-----KIQRLKLCRALQTERNDLNKRVODLSA 474
DB 453 ETLKRENKNOQEISSLTEQIAEGGKRIHELEKIKQIQEOKSLSLQALBEAEA 506

RESULT 41
US-10-408-765A-1168
; Sequence 1168, Application US/10/408765A
; Publication NO. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 66088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1168
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1168

Query Match
Best Local Similarity 8.8%; Score 237; DB 16; Length 1938;
Matches 111; Conservative 113; Mismatches 197; Indels 98; Gaps 20;

QY 19 SQAPPAVEAGGSGSQAPKPGCAQRTAQSALDVSE-----LSR 61
DB 1092 SOLQARIDDEQVHSLQFOKIKELQAR-----FELEERIEAEHTLRAKIEKQSDLAR 1145
QY 62 QLEDILSTYCVNNQGGPGGAGGAPAEPEDAE---KSRTYVARNGEPPTPVVYGEKE 118
DB 1146 ELSEI-----SERLEBASGATSQIENWKKEAFQWRDLEZATLQHEATAATLRKKQ 1200
QY 119 PSKGDPTTEIRQSDVGRDHRPQEKKAKG-LGKEITLLMOTINTLSTPEKLAALC 177
DB 1201 -----ADSVAEELGEQIDNLRVQKLEKEKSELKMEIDDMASNIKSKSNIRTC 1253
QY 178 KYAELLEHEHNSQKMKLLQKQSQLV-----QEKHLRGEHSAVLARSKLSLSCRELQ 233
DB 1254 RIVEDQPE-----IAKDEQQTQLHLDNMQARLOTQNGELSHRVEKESLSLSQLT 1306
QY 234 RHNSLKE-EGVQARAREEERKEVTSK-FQVTLNDIQLQMEQHNERNKLRQENNELA 290
DB 1307 KSKQALTQQLSELKQWEEETKAKNAHAHALQSSRHDCDLLEQYBEE---QKAKAELQ 1362
QY 291 ERLKKI---IQYSLRERHIDKVPKHDLQO---CLVDKALQQAOS-----MLKEA 335
DB 1363 RALSKANSEVAQWTKYE-TDAIORTLEBAKKLAQRLQEAETETANSKASLEKT 1421
QY 336 BERHQREKDFLLKSAVSSQRMCKMQQETHLQKQALYTEKFEFPQNTL-----SK 387
DB 1422 KQRLQGEVEDLRDLERSHTACATLDKKQRNFDKVLAEWKQKLDQESQAELEAAQKESRS 1481
QY 388 SSEVFT---TFKQ---EMEKTKKIKKLEKTTMYRSWESSNKALLEMAEKTVRDKEL 441
DB 1482 STELFQKNRANRYEVVDLETLRRENKLOBEISDLTEQIATGKNLQEAETKLVQEQEK 1541
QY 442 EQLQVKIQRL-----KLQRA---IQTERNDLNKRV 469
DB 1542 SDLQVALEEVGSLSEHSESKILRVQLELSQVKSLEDRKV 1580

RESULT 42
US-10-336-472-16
; Sequence 16, Application US/10336472
; Publication NO. US20040043929A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.
; APPLICANT: Ballinger, Robert A.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Bergths, Constance
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E.

```

; Sequence 6752, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6752
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6752

Query Match 1.3%; Score 7; DB 4; Length 474;
Best Local Similarity 100.0%; Pred. No. 3.6e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 BIRQSD 134
Db 386 BIRQSD 392

RESULT 47
US-08-962-859A-2
; Sequence 2, Application US/08962859A
; Patent No. 6127345
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin K. R.
; TITLE OF INVENTION: No. 6127345el Glucose 6-Phosphate
; FILE REFERENCE: Dehydrogenase Gene
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/962,859A
; FILING DATE: 03-NOV-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER: 60/035,072
; FILING DATE: 21-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson, Arthur E
; REGISTRATION NUMBER: 34,354
; REFERENCE/DOCKET NUMBER: GM50001
; TELEPHONE: 609/520-3254
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 495 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-962-859A-2

Query Match 1.3%; Score 7; DB 3; Length 495;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 48 QSCALRD 54
Db 228 QSCALRD 234

RESULT 48
US-09-252-991A-20533
; Sequence 20533, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20533
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20533

Query Match 1.3%; Score 7; DB 4; Length 526;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 AELLEEH 187
Db 122 AELLEEH 128

RESULT 49
US-08-285-440-5
; Sequence 5, Application US/08285440
; Patent No. 5532337
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/285,440
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
PUBLICATION RESIDUES IN SEQ ID NO:
US-08-285-440-5

Query Match 1.3%; Score 7; DB 1; Length 532;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 248 REBEKR 254
Db 335 REBEKR 341

RESULT 50
US-08-630-349-5
Sequence 5, Application US/08630349
Patent No. 5739008
GENERAL INFORMATION:
APPLICANT: Ken'ichiro HAYASHI et al.
TITLE OF INVENTION: POLYPEPTIDE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESS: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,349
FILING DATE: April 10, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/285,440
FILING DATE: August 4, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/858,947
FILING DATE: March 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-630-349-5

Query Match 1.3%; Score 7; DB 1; Length 532;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 248 REBEKR 254
Db 335 REBEKR 341

RESULT 51
US-09-138-277C-3
Sequence 3, Application US/09138277C
Patent No. 6426403
GENERAL INFORMATION:
APPLICANT: NAKATA, MOTOMI
APPLICANT: NAKANO, HIROYASU
APPLICANT: YAGITA, HIDEO
APPLICANT: OKUMURA, KO
TITLE OF INVENTION: TRAP FAMILY MOLECULES, POLYNUCLEOTIDES ENCODING THEM,
TITLE OF INVENTION: AND ANTIBODIES AGAINST THEM
FILE REFERENCE: 007898-0255515
CURRENT APPLICATION NUMBER: US/09/138,277C
CURRENT FILING DATE: 1998-08-18
PRIOR APPLICATION NUMBER: PCT/JP97/00512
PRIOR FILING DATE: 1997-02-24
PRIOR APPLICATION NUMBER: JP 34674/1996
PRIOR FILING DATE: 1996-02-22
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 3
LENGTH: 557
TYPE: PRT
ORGANISM: Homo sapiens
US-09-138-277C-3

Query Match 1.3%; Score 7; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 406 IKLEKE 412
Db 285 IKLEKE 291

RESULT 52
US-08-285-440-6
Sequence 6, Application US/08285440
Patent No. 5532337
GENERAL INFORMATION:

APPLICANT: Ken'ichiro HAYASHI et al.
TITLE OF INVENTION: POLYPEPTIDE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: DisplayWrite
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/285,440
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/859,947
FILING DATE: March 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-6
Query Match 1.3%; Score 7; DB 1; Length 558;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
QY 248 REEEKR 254
Db 361 REEEKR 367
RESULT 53
US-08-630-349-6
; Sequence 6, Application US/08630349
; Patent No. 5739008
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,349
; FILING DATE: April 10, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/285,440
; FILING DATE: August 4, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 558 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-630-349-6
Query Match 1.3%; Score 7; DB 1; Length 558;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
QY 248 REEEKR 254
Db 361 REEEKR 367
RESULT 54
US-09-499-522-18
; Sequence 18, Application US/09499522
; Patent No. 6479238
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; TITLE OF INVENTION: POLYMORPHIC MARKERS OF THE LSR GENE
; FILE REFERENCE: GENSET.053AUS

; CURRENT APPLICATION NUMBER: US/09/499,522
; CURRENT FILING DATE: 2000-02-10
; EARLIER APPLICATION NUMBER: US 60/119,592
; EARLIER FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: US 60/144,784
; EARLIER FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent.pm
; SEQ ID NO 18
; LENGTH: 581
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 295
; OTHER INFORMATION: 9-7-325 : polymorphic amino acid Ser or Asn
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 352
; OTHER INFORMATION: 9-9-246 : polymorphic amino acid Pro or Arg
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 451
; OTHER INFORMATION: LSRX9f13-BM : polymorphic amino acid deletion of Arg
US-09-499-522-18

Query Match 1.3%; Score 7; DB 4; Length 581;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 SAGGQGS 479
|||
Db 297 SAGGQGS 303

RESULT 55
US-09-269-939A-12
; Sequence 12, Application US/09269939A
; Patent No. 6635431
; GENERAL INFORMATION:
; APPLICANT: Bihaun, Bernard
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Yen-Potin, Frances
; TITLE OF INVENTION: LSR Receptor, Its Activity, Its Cloning and Its Application To
; TITLE OF INVENTION: The Diagnosis Prevention And/or Treatment of Obesity and
; TITLE OF INVENTION: Related Risks or Complications
; FILE REFERENCE: GENSET.035APC
; CURRENT APPLICATION NUMBER: US/09/269,939A
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: FR 97/10088
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: FR 98/05032
; PRIOR FILING DATE: 1998-04-22
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent.pm
; SEQ ID NO 12
; LENGTH: 591
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-269-939A-12

Query Match 1.3%; Score 7; DB 4; Length 581;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 SAGGQGS 479
|||
Db 297 SAGGQGS 303

RESULT 56
US-09-252-991A-27626
; Sequence 27626, Application US/09252991A

; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27626
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27626

Query Match 1.3%; Score 7; DB 4; Length 582;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 155 EITLLMQ 161
|||
Db 355 EITLLMQ 361

RESULT 57
US-09-499-522-16
; Sequence 16, Application US/09499522
; Patent No. 6479238
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihaun, Bernard
; TITLE OF INVENTION: POLYMORPHIC MARKERS OF THE LSR GENE
; FILE REFERENCE: GENSET.053AUS
; CURRENT APPLICATION NUMBER: US/09/499,522
; CURRENT FILING DATE: 2000-02-10
; EARLIER APPLICATION NUMBER: US 60/119,592
; EARLIER FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: US 60/144,784
; EARLIER FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent.pm
; SEQ ID NO 16
; LENGTH: 630
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 344
; OTHER INFORMATION: 9-7-325 : polymorphic amino acid Ser or Asn
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 401
; OTHER INFORMATION: 9-9-246 : polymorphic amino acid Pro or Arg
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 500
; OTHER INFORMATION: LSRX9f13-BM : polymorphic amino acid deletion of Arg
US-09-499-522-16

Query Match 1.3%; Score 7; DB 4; Length 630;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 SAGGQGS 479
|||
Db 346 SAGGQGS 352

```

RESULT 58
US-09-269-939A-10
; Sequence 10, Application US/09269939A
; Patent No. 6635431
; GENERAL INFORMATION:
; APPLICANT: Bihain, Bernard
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: LSR Receptor, Its Activity, Its Cloning and Its Application To
; TITLE OF INVENTION: The Diagnosis Prevention And/or Treatment of Obesity and
; TITLE OF INVENTION: Related Risks or Complications
; FILE REFERENCE: GENSET.035APC
; CURRENT APPLICATION NUMBER: US/09/269,939A
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: FR 97/10088
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: FR 98/05032
; PRIOR FILING DATE: 1998-04-22
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent.pm
; SEQ ID NO 10
; LENGTH: 630
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-269-939A-10

Query Match 1.3%; Score 7; DB 4; Length 630;
Best Local Similarity 100.0%; Pred. No. 4.6e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy 473 SAGGQGS 479
Db 346 SAGGQGS 352

RESULT 59
US-09-499-522-14
; Sequence 14, Application US/09499522
; Patent No. 6479238
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; TITLE OF INVENTION: POLYMORPHIC MARKERS OF THE LSR GENE
; FILE REFERENCE: GENSET.053AUS
; CURRENT APPLICATION NUMBER: US/09/499,522
; CURRENT FILING DATE: 2000-02-10
; EARLIER APPLICATION NUMBER: US 60/119,592
; EARLIER FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: US 60/144,784
; EARLIER FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent.pm
; SEQ ID NO 14
; LENGTH: 649
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 363
; OTHER INFORMATION: 9-7-325 : polymorphic amino acid Ser or Asn
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 420
; OTHER INFORMATION: 9-9-246 : polymorphic amino acid Pro or Arg
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 519
; OTHER INFORMATION: LSRX9f13-BM : polymorphic amino acid deletion of Arg
US-09-499-522-14

Query Match 1.3%; Score 7; DB 4; Length 649;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 SAGGQGS 479
Db 365 SAGGQGS 371

RESULT 60
US-09-269-939A-8
; Sequence 8, Application US/09269939A
; Patent No. 6635431
; GENERAL INFORMATION:
; APPLICANT: Bihain, Bernard
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Yen-Potin, Frances
; TITLE OF INVENTION: LSR Receptor, Its Activity, Its Cloning and Its Application To
; TITLE OF INVENTION: The Diagnosis Prevention And/or Treatment of Obesity and
; TITLE OF INVENTION: Related Risks or Complications
; FILE REFERENCE: GENSET.035APC
; CURRENT APPLICATION NUMBER: US/09/269,939A
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: FR 97/10088
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: FR 98/05032
; PRIOR FILING DATE: 1998-04-22
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent.pm
; SEQ ID NO 8
; LENGTH: 649
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Misc_Feature
; LOCATION: 386
; OTHER INFORMATION: Potential deletion of a Glu
; NAME/KEY: Misc_Feature
; LOCATION: 518
; OTHER INFORMATION: Potential insertion of a Arg
US-09-269-939A-8

Query Match 1.3%; Score 7; DB 4; Length 649;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 SAGGQGS 479
Db 365 SAGGQGS 371

RESULT 61
US-09-543-681A-5392
; Sequence 5392, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5392
; LENGTH: 712
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5392

Query Match 1.3%; Score 7; DB 4; Length 712;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 46 TAQSAL 52
|||||
DB 373 TAQSAL 379

RESULT 62
US-09-562-737-47
; Sequence 47, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 47
; LENGTH: 1024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-47

Query Match 1.3%; Score 7; DB 4; Length 1024;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 KDFLLKE 349
|||||
DB 263 KDFLLKE 269

RESULT 63
US-09-252-991A-18387
; Sequence 18387, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107195.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18387
; LENGTH: 1075
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18387

Query Match 1.3%; Score 7; DB 4; Length 1075;
Best Local Similarity 100.0%; Pred. No. 7.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 QAAPAVE 26
|||||
DB 999 QAAPAVE 1005

RESULT 64
US-08-131-365B-54
; Sequence 54, Application US/08131365B
; Patent No. 5527690
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.

; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/131,365B
; APPLICATION NUMBER: US/08/131,365B
; FILING DATE: 01-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1141
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-131-365B-54

Query Match 1.3%; Score 7; DB 1; Length 1141;
Best Local Similarity 100.0%; Pred. No. 7.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 KLRQENM 287
|||||
DB 386 KLRQENM 392

RESULT 65
US-08-668-123-54
; Sequence 54, Application US/08668123
; Patent No. 5891631
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,123
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/131,365
; FILING DATE: 01-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1141
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-668-123-54

Query Match 1.3%; Score 7; DB 2; Length 1141;
Best Local Similarity 100.0%; Pred. No. 7.5e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 281 KLRQNM 287
| | | | |
Db 386 KLRQNM 392

RESULT 66
US-09-976-594-736
; Sequence 736, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 736
; LENGTH: 1530
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 3151579CD1
US-09-976-594-736

Query Match 1.3%; Score 7; DB 4; Length 1530;
Best Local Similarity 100.0%; Pred. No. 9.5e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 439 KELEGLQ 445
| | | | |
Db 1316 KELEGLQ 1322

RESULT 67
US-08-287-959-1
; Sequence 1, Application US/08287959
; Patent No. 5639651
; GENERAL INFORMATION:
; APPLICANT: Weissbach, Lawrence
; APPLICANT: Bernards, Andre
; APPLICANT: Settleman, Jeffrey

TITLE OF INVENTION: GAP-RELATED GENE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/287,959
; FILING DATE: August 9, 1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul C.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/181001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1657 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-287-959-1

Query Match 1.3%; Score 7; DB 1; Length 1657;
Best Local Similarity 100.0%; Pred. No. 1e+03; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 471 DLSAGGQ 477
| | | | |
Db 1210 DLSAGGQ 1216

RESULT 68
US-09-252-991A-23568
; Sequence 23568, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23568
; LENGTH: 3340
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23568

Query Match 1.3%; Score 7; DB 4; Length 3340;
Best Local Similarity 100.0%; Pred. No. 1.8e+03; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 365 THLKQQL 371
| | | | |
Db 2522 THLKQQL 2528

RESULT 69
US-08-310-912A-138
; Sequence 138, Application US/08310912A
; Patent No. 5981730
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND DETECTION
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/310.912A
; FILING DATE: September 22, 1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/227,360
; FILING DATE: April 13, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/254001
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; MOLECULE TYPE: protein
US-08-310-912A-138

Query Match 1.1%; Score 6; DB 2; Length 6;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 SKLESL 228
Db 1 SKLESL 6

RESULT 71
PCT-US95-04589-138
; Sequence 138, Application PC/TUS9504589
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE AND USBS THEREOF
; NUMBER OF SEQUENCES: 201
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04589
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,360
; FILING DATE: 13-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

US-08-310-912A-138
; Sequence 138, Application US/08310912A
; Patent No. 5981730
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND DETECTION
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/310.912A
; FILING DATE: September 22, 1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/227,360
; FILING DATE: April 13, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/254001
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-310-912A-138

Query Match 1.1%; Score 6; DB 2; Length 6;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 SKLESL 228
Db 1 SKLESL 6

RESULT 70
US-09-301-085-138
; Sequence 138, Application US/09301085
; Patent No. 6262248
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.

MOLECULE TYPE: protein

PCT-US95-04589-138

Query Match

1.1%; Score 6; DB 5; Length 6;

Best Local Similarity 100.0%; Pred. No. 3e+05;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 SKLESL 228

|||||

Db 1 SKLESL 6

RESULT 72

US-09-989-789-1988

Sequence 1988, Application US/09989789

Patent No. 8588746

GENERAL INFORMATION:

APPLICANT: LIU, Qiang

TITLE OF INVENTION: POSITION DEPENDENT RECOGNITION OF GNN NUCLEOTIDE

TITLE OF INVENTION: TRIPLETS BY ZINC FINGERS

FILE REFERENCE: 8325-0011.20 / S11-US2

CURRENT APPLICATION NUMBER: US/09/989,789

CURRENT FILING DATE: 2002-03-25

NUMBER OF SEQ ID NOS: 4085

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 1988

LENGTH: 7

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: example ZFP

US-09-989-789-1988

Query Match

1.1%; Score 6; DB 4; Length 7;

Best Local Similarity 100.0%; Pred. No. 3e+05;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 QSGALR 53

|||||

Db 1 QSGALR 6

RESULT 73

US-08-339-152A-1

Sequence 1, Application US/08339152A

Patent No. 5643726

GENERAL INFORMATION:

APPLICANT: Tanzi, Rudolph E.

APPLICANT: Kovacs, Dora M.

TITLE OF INVENTION: Methods For Modulating Transcription

TITLE OF INVENTION: From The Amyloid -Protein Precursor (APP) Promoter

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.

STREET: 1100 New York Ave., NW, Suite 600

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/339,152A

FILING DATE: 10-NOV-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Steffe, Eric K.

REGISTRATION NUMBER: 36,688

REFERENCE/DOCKET NUMBER: 0609.4120000

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-2600

TELEFAX: 202-371-2540

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 9 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-339-152A-1

Query Match

1.1%; Score 6; DB 1; Length 9;

Best Local Similarity 100.0%; Pred. No. 3e+05;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 RELORH 235

|||||

Db 4 RELORH 9

RESULT 74

US-08-007-999B-1

Sequence 1, Application US/08007999B

Patent No. 5851787

GENERAL INFORMATION:

APPLICANT: Wasco, Wilma

APPLICANT: Bupp, Keith

APPLICANT: Magendantz, Margaret

APPLICANT: Tanzi, Rudolph

APPLICANT: Solomon, Frank

TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES

TITLE OF INVENTION: THEREOF

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox, P.L.L.C.

STREET: 1100 New York Ave., NW

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/007,999B

FILING DATE: 21-JAN-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/872,642

FILING DATE: 20-APR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/930,022

FILING DATE: 17-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: Townsend, G. Kevin

REGISTRATION NUMBER: 34,033

REFERENCE/DOCKET NUMBER: 0609.3520002/JAG/GKT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)371-2571

TELEFAX: (202)371-2540

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 9 amino acids

TYPE: amino acid

STRANDEDNESS: both

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-007-999B-1

Query Match

1.1%; Score 6; DB 2; Length 9;

Best Local Similarity 100.0%; Pred. No. 3e+05;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 RELQKH 235
|||||
Db 4 RELQKH 9

RESULT 75

US-08-689-276A-1
; Sequence 1, Application US/08689276A
; Patent No. 5891991
; GENERAL INFORMATION:
; APPLICANT: Masco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendanz, Margaret
; APPLICANT: Tanzi, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,276A
; FILING DATE: 06-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/007,999
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: JORGE A. GOLDSTEIN
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 0609.3520003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2571
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-689-276A-1

Query Match 1.1%; Score 6; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 3e+05; 0; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 RELQKH 235
|||||
Db 4 RELQKH 9

RESULT 76

US-08-159-339A-830
; Sequence 830, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.

; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/925,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 018623-005030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX:
; INFORMATION FOR SEQ ID NO: 830:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-159-339A-830

Query Match 1.1%; Score 6; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; 0; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GPEGQAQ 15
|||||
Db 2 GPEGQAQ 7

RESULT 77

US-08-836-075A-109
; Sequence 109, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; TITLE OF INVENTION: AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433


```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/BP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-836-075A-109

Query Match 1.1%; Score 6; DB 3; Length 11;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 486 ERRPEG 491
DB 1 ERRPEG 6

RESULT 78
US-08-747-599A-16
; Sequence 16, Application US/08747599A
; Patent No. 6214795
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Howard et al.
; TITLE OF INVENTION: Peptide Compounds Useful for Modulating
; TITLE OF INVENTION: FGF Receptor Activity
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/747,599A
; FILING DATE: 12-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kara, Catherine J.
; REGISTRATION NUMBER: 41,106
; REFERENCE/DOCKET NUMBER: PPI-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids

;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; US-08-747-599A-16

Query Match 1.1%; Score 6; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LALYTE 376
DB 4 LALYTE 9

RESULT 79
US-08-979-608A-29
; Sequence 29, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-No. 6355451-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 29:
; US-08-979-608A-29

Query Match 1.1%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 EPEDAE 95
DB 10 EPEDAE 15
```


Db 13 EEEKR 18

RESULT 87

US-09-834-759-544

Sequence 544, Application US/09834759

Patent No. 6680197

GENERAL INFORMATION:

APPLICANT: Jiang, Yugu

APPLICANT: Dillon, Davin C.

APPLICANT: Mitcham, Jennifer L.

APPLICANT: Xu, Jiangchun

APPLICANT: Harlocker, Susan L.

APPLICANT: Hepler, William T.

APPLICANT: Henderson, Robert A.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.470C9

CURRENT FILING DATE: 2001-04-13

CURRENT APPLICATION NUMBER: US/09/834,759

NUMBER OF SEQ ID NOS: 547

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 544

LENGTH: 20

TYPE: PRT

ORGANISM: Homo sapiens

US-09-834-759-544

Query Match 1.1%; Score 6; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 2.2e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0;

QY 316 LQQQLV 321

Db 5 LQQQLV 10

RESULT 88

US-09-834-759-543

Sequence 543, Application US/09834759

Patent No. 6680197

GENERAL INFORMATION:

APPLICANT: Jiang, Yugu

APPLICANT: Dillon, Davin C.

APPLICANT: Mitcham, Jennifer L.

APPLICANT: Xu, Jiangchun

APPLICANT: Harlocker, Susan L.

APPLICANT: Hepler, William T.

APPLICANT: Henderson, Robert A.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.470C9

CURRENT FILING DATE: 2001-04-13

CURRENT APPLICATION NUMBER: US/09/834,759

NUMBER OF SEQ ID NOS: 547

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 543

LENGTH: 21

TYPE: PRT

ORGANISM: Homo sapiens

US-09-834-759-543

Query Match 1.1%; Score 6; DB 4; Length 21;

Best Local Similarity 100.0%; Pred. No. 2.3e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0;

QY 316 LQQQLV 321

Db 14 LQQQLV 19

RESULT 89

US-08-851-843A-113

Sequence 113, Application US/08851843A

Patent No. 6093809

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.

APPLICANT: Lingner, Joachim

APPLICANT: Nakamura, Toru

APPLICANT: Chapman, Karen B.

APPLICANT: Morin, Gregg B.

APPLICANT: Harley, Calvin

APPLICANT: Andrews, William H.

TITLE OF INVENTION: No. 6093809el Telomerase

NUMBER OF SEQUENCES: 225

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94111

COMPUTER READABLE FORM: disk

MEDIUM TYPE: Floppy

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/851,843A

FILING DATE: 06-MAY-1997

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/846,017

FILING DATE: 25-APR-1997

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/844,419

FILING DATE: 18-APR-1997

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/724,643

FILING DATE: 01-OCT-1996

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph T.

REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002930US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 113:

SEQUENCE CHARACTERISTICS:

LENGTH: 23 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-851-843A-113

Query Match 1.1%; Score 6; DB 3; Length 23;

Best Local Similarity 100.0%; Pred. No. 2.5e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0;

QY 52 LRDVSE 57

Db 9 LRDVSE 14

RESULT 90

US-08-974-549A-233

Sequence 233, Application US/08974549A

Patent No. 6166178

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.

APPLICANT: Lingner, Joachim

APPLICANT: Nakamura, Toru

APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3634
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0300
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 233:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-974-549A-233

Query Match 1.1%; Score 6; DB 3; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0;

QY 52 LRDVSE 57

DB 9 LRDVSE 14
RESULT 91
US-08-854-050-113
Sequence 113, Application US/08854050
Patent No. 6261836
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6261836el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,050
FILING DATE: 09-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-854-050-113

Query Match 1.1%; Score 6; DB 3; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0;

QY 52 LRDVSE 57

DB 9 LRDVSE 14

RESULT 92

US-09-430-323-113
; Sequence 113, Application US/09430323
; Patent No. 6309867

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.
; Lingner, Joachim
; Nakamura, Toru
; Chapman, Karen B.
; Morin, Gregg B.
; Harley, Calvin
; Andrews, William H.
TITLE OF INVENTION: NO. 6309867el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/430,323
FILING DATE: 29-Oct-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
CLASSIFICATION: US 08/851,843
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996

ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0300
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 113:

SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 113:

US-09-430-323-113

Query Match 1.1%; Score 6; DB 4; Length 23;

Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 LRDVSE 57

DB 9 LRDVSE 14

RESULT 93

US-08-912-951-114

; Sequence 114, Application US/08912951

; Patent No. 6475789

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.
; Lingner, Joachim
; Nakamura, Toru
; Chapman, Karen B.
; Morin, Gregg B.
; Harley, Calvin
; Andrews, William H.
TITLE OF INVENTION: HUMAN TELOMERASE CATALYTIC SUBUNIT: DIAGNOSTIC AND
THERAPEUTIC METHODS
NUMBER OF SEQUENCES: 335
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/912,951
FILING DATE: 14-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph T.

REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002600US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0300

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 114:

SEQUENCE CHARACTERISTICS:

LENGTH: 23 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-912-951-114

Query Match 1.1%; Score 6; DB 4; Length 23;

Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 LRDVSE 57

DB 9 LRDVSE 14

RESULT 94

US-09-402-181B-233
; Sequence 233, Application US/09402181B
; Patent No. 6610839
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; Lingner, Joachim
; Nakamura, Toru
; Chapman, Karen B.
; Morin, Gregg B.
; Harley, Calvin B.
; Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 633
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/402,181B
; FILING DATE: 29-Sep-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: US 08/912,951
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: WO PCT/US97/17885
; FILING DATE: 01-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Aussenhus, Scott L.
; REGISTRATION NUMBER: 42,271
; REFERENCE/DOCKET NUMBER: 015389-002620US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 233:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 233:
US-09-402-181B-233
Query Match 1.1%; Score 6; DB 4; Length 23;
Best local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
52 LRDVSE 57
| | | | |
9 LRDVSE 14
Db

RESULT 95
US-09-721-456-233
; Sequence 233, Application US/09721456
; Patent No. 6617110
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; Lingner, Joachim
; Nakamura, Toru
; Chapman, Karen B.
; Morin, Gregg B.
; Harley, Calvin B.
; Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/721,456
; FILING DATE: 22-No. 6617110-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/974,549A
; FILING DATE: 19-NOV-1997
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: US 08/912,951
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: US 08/915,503
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: WO PCT/US97/17618
; FILING DATE: 01-OCT-1997
; APPLICATION NUMBER: WO PCT/US97/17885
; FILING DATE: 01-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph Ted
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002610US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 233:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 233:
US-09-721-456-233
Query Match 1.1%; Score 6; DB 4; Length 23;
Best local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 LRQVSE 57
|||||
Db 9 LRQVSE 14

RESULT 96
US-08-507-598-4
; Sequence 4, Application US/08507598
; Patent No. 5934188
; GENERAL INFORMATION:
; APPLICANT: HARADA, SHUN-ICHI
; APPLICANT: SAMPATH, T. K.
; APPLICANT: RODAN, GIDEON A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
; ADDRESSEE: THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/507,598
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)-248-7000
; TELEFAX: (617)-248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..25
; OTHER INFORMATION: /note= "Conserved domain of human
; US-08-507-598-4

Query Match 1.1%; Score 6; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEKR 254
|||||
Db 8 EEEKR 13

RESULT 97
US-08-507-750-4
; Sequence 4, Application US/08507750
; Patent No. 5932716
; GENERAL INFORMATION:
; APPLICANT: SAMPATH, T. K.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING
; MORPHOGEN ANALOGS

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
ADDRESSEE: THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/507,750
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER, EDMUND R.
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: CRP-116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)-248-7000
TELEFAX: (617)-248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..25
OTHER INFORMATION: /note= "Conserved domain of human
US-08-507-750-4

Query Match 1.1%; Score 6; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEKR 254
|||||
Db 8 EEEKR 13

RESULT 98
US-08-764-522A-4
; Sequence 4, Application US/08764522A
; Patent No. 6090544
; GENERAL INFORMATION:
; APPLICANT: HARADA, SHUN-ICHI
; APPLICANT: SAMPATH, T. K.
; APPLICANT: RODAN, GIDEON A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING
; MORPHOGEN ANALOGS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,522A

;; FILING DATE: 435
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: VITO, CHRISTINE C.
;; REGISTRATION NUMBER: 39,061
;; REFERENCE/DOCKET NUMBER: CRP-126
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617)-248-7000
;; TELEFAX: (617)-248-7100
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 25 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: Peptide
;; FEATURE:
;; NAME/KEY: Peptide
;; LOCATION: 1..25
;; OTHER INFORMATION: /note= "Conserved domain of human
;; OTHER INFORMATION: c-fos"
;; US-08-764-522A-4

Query Match 1.1%; Score 6; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEER 254
DB 8 EEEER 13

RESULT 99
US-08-764-528-4
;; Sequence 4, Application US/08764528
;; Patent No. 6103491
;; GENERAL INFORMATION:
;; APPLICANT: SAMPATH, K. T.
;; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING
;; TITLE OF INVENTION: MORPHOGEN ANALOGS
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
;; STREET: 45 SOUTH STREET
;; CITY: HOPKINTON
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 01748
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/764,528
;; FILING DATE:
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: VITO, CHRISTINE C.
;; REGISTRATION NUMBER: 39,061
;; REFERENCE/DOCKET NUMBER: CRP-127
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617)-248-7000
;; TELEFAX: (617)-248-7100
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 25 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; FEATURE:
;; NAME/KEY: Peptide

;; LOCATION: 1..25
;; OTHER INFORMATION: /note= "Conserved domain of human
;; OTHER INFORMATION: c-fos"
;; US-08-764-528-4
Query Match 1.1%; Score 6; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEER 254
DB 8 EEEER 13

RESULT 100
US-08-872-859-4
;; Sequence 4, Application US/08872859
;; Patent No. 6110460
;; GENERAL INFORMATION:
;; APPLICANT: SAMPATH, T. K.
;; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING
;; TITLE OF INVENTION: MORPHOGEN ANALOGS
;; NUMBER OF SEQUENCES: 4
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
;; ADDRESSEE: THIBEAULT
;; STREET: 53 STATE STREET
;; CITY: BOSTON
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02109
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/872,859
;; FILING DATE: 11-JUN-1997
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/507,750
;; FILING DATE: 26-JUL-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: PITCHER, EDMUND R.
;; REGISTRATION NUMBER: 27,829
;; REFERENCE/DOCKET NUMBER: CRP-116
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617)-248-7000
;; TELEFAX: (617)-248-7100
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 25 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; FEATURE:
;; NAME/KEY: Peptide
;; LOCATION: 1..25
;; OTHER INFORMATION: /note= "Conserved domain of human
;; OTHER INFORMATION: c-fos"
;; US-08-872-859-4

Query Match 1.1%; Score 6; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEER 254
DB 8 EEEER 13

Mon Jun 7 17:24:59 2004

us-10-023-529-8.oli.ra1

Page 36

Search completed: June 7, 2004, 14:47:22
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:45:27 ; Search time 41 Seconds
(without alignments)
3636.817 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 530

Sequence: 1 KSSPGPAGPCAQRPSQ.....APSTEASGQTGQEPSTARA 530

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1155919 seqs, 281338677 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 300 summaries

Database : Published Applications AA:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	530	100.0	530	9	US-09-962-055-8
2	530	100.0	530	9	US-09-976-740-8
3	530	100.0	530	12	US-10-671-242-8
4	530	100.0	530	13	US-10-023-529-8
5	530	100.0	530	13	US-10-023-523-8
6	530	100.0	530	15	US-10-616-187-8
7	429	80.9	546	9	US-09-976-740-44
8	429	80.9	546	12	US-10-671-242-44
9	429	80.9	546	13	US-10-023-529-44
10	429	80.9	546	13	US-10-023-523-44
11	429	80.9	546	15	US-10-616-187-44
12	359	67.7	510	12	US-10-276-774-2134
13	172	32.5	557	9	US-09-962-055-5
14	172	32.5	557	9	US-09-976-740-5
15	172	32.5	557	12	US-10-671-242-5

16	172	32.5	557	13	US-10-023-529-5	Sequence 5, Appli
17	172	32.5	557	13	US-10-023-523-5	Sequence 5, Appli
18	172	32.5	557	15	US-10-616-187-5	Sequence 5, Appli
19	17	3.2	85	15	US-10-264-049-3413	Sequence 3413, Ap
20	11	2.1	436	12	US-10-424-599-207434	Sequence 207434,
21	9	1.7	99	9	US-09-764-869-1082	Sequence 1082, Ap
22	9	1.7	99	14	US-10-091-504-1082	Sequence 1082, Ap
23	9	1.7	99	15	US-10-227-577-1082	Sequence 112, App
24	9	1.7	594	15	US-09-801-368-112	Sequence 1762, Ap
25	9	1.7	594	15	US-10-369-493-1762	Sequence 72115, A
26	9	1.7	637	12	US-10-382-122A-72115	Sequence 130, App
27	8	1.5	25	16	US-10-443-622-130	Sequence 48446, A
28	8	1.5	106	12	US-10-425-114-48446	Sequence 272074
29	8	1.5	121	12	US-10-424-599-272074	Sequence 132, App
30	8	1.5	156	16	US-10-443-622-122	Sequence 316, App
31	8	1.5	229	14	US-10-080-170-316	Sequence 12433, A
32	8	1.5	413	14	US-10-156-761-12433	Sequence 21212, A
33	8	1.5	447	15	US-10-369-493-21212	Sequence 7865, Ap
34	8	1.5	450	12	US-10-335-977-7865	Sequence 7866, Ap
35	8	1.5	451	12	US-10-335-977-7866	Sequence 6430, Ap
36	8	1.5	484	9	US-09-738-626-6430	Sequence 23425, A
37	8	1.5	525	15	US-10-389-493-23425	Sequence 41, Appl
38	7	1.3	7	9	US-09-962-055-41	Sequence 41, Appl
39	7	1.3	7	9	US-09-976-740-41	Sequence 41, Appl
40	7	1.3	7	12	US-10-671-242-41	Sequence 41, Appl
41	7	1.3	7	13	US-10-023-529-41	Sequence 41, Appl
42	7	1.3	7	13	US-10-023-523-41	Sequence 41, Appl
43	7	1.3	7	15	US-10-616-187-41	Sequence 2071, Ap
44	7	1.3	18	14	US-10-225-567A-2071	Sequence 348, App
45	7	1.3	27	14	US-10-195-730-348	Sequence 36055, A
46	7	1.3	38	9	US-09-864-761-36055	Sequence 31, Appl
47	7	1.3	44	10	US-09-491-614-31	Sequence 335, App
48	7	1.3	45	10	US-09-986-480-335	Sequence 355, App
49	7	1.3	45	10	US-09-892-877-355	Sequence 359, App
50	7	1.3	45	10	US-09-948-783-359	Sequence 14654, A
51	7	1.3	55	12	US-10-424-599-154654	Sequence 246225,
52	7	1.3	78	12	US-10-424-599-246225	Sequence 68147, A
53	7	1.3	80	12	US-10-425-114-68147	Sequence 1230, Ap
54	7	1.3	82	9	US-09-796-692-1230	Sequence 1230, Ap
55	7	1.3	82	14	US-10-040-862-1230	Sequence 1230, Ap
56	7	1.3	82	15	US-10-057-475B-1230	Sequence 1236, Ap
57	7	1.3	82	15	US-10-154-884B-1230	Sequence 1236, Ap
58	7	1.3	89	12	US-10-424-599-197528	Sequence 197528,
59	7	1.3	91	9	US-09-796-692-1236	Sequence 1236, Ap
60	7	1.3	91	9	US-09-796-692-1693	Sequence 1693, Ap
61	7	1.3	91	14	US-10-040-862-1236	Sequence 1236, Ap
62	7	1.3	91	14	US-10-040-862-1693	Sequence 1236, Ap
63	7	1.3	91	15	US-10-057-475B-1236	Sequence 1693, Ap
64	7	1.3	91	15	US-10-057-475B-1693	Sequence 1236, Ap
65	7	1.3	91	15	US-10-154-884B-1236	Sequence 1693, Ap
66	7	1.3	91	15	US-10-154-884B-1693	Sequence 39633, A
67	7	1.3	95	9	US-09-864-761-39633	Sequence 56, Appl
68	7	1.3	99	12	US-09-804-014A-56	Sequence 26203,
69	7	1.3	100	12	US-10-424-599-266203	Sequence 153810,
70	7	1.3	101	12	US-10-424-599-153810	Sequence 4293, Ap
71	7	1.3	106	15	US-10-264-049-4293	Sequence 345, App
72	7	1.3	110	14	US-10-195-730-345	Sequence 50441, A
73	7	1.3	113	12	US-10-425-114-50441	Sequence 165582,
74	7	1.3	114	12	US-10-424-599-165582	Sequence 196307,
75	7	1.3	115	14	US-10-424-599-196307	Sequence 334, App
76	7	1.3	115	14	US-10-043-487-334	Sequence 4955, Ap
77	7	1.3	117	12	US-10-335-977-4955	Sequence 241987,
78	7	1.3	119	12	US-10-424-599-241987	Sequence 121, App
79	7	1.3	122	16	US-10-443-622-121	Sequence 150351,
80	7	1.3	124	12	US-10-424-599-150351	Sequence 19, Appl
81	7	1.3	127	8	US-08-849-303-19	Sequence 3437, Ap
82	7	1.3	130	15	US-10-108-260A-3437	Sequence 4956, Ap
83	7	1.3	133	12	US-10-335-977-4956	Sequence 4957, Ap
84	7	1.3	133	12	US-10-335-977-4957	Sequence 4958, Ap
85	7	1.3	137	12	US-10-335-977-4958	Sequence 130, App
86	7	1.3	138	10	US-09-847-208-130	Sequence 128, App
87	7	1.3	139	10	US-09-847-208-128	Sequence 69680, A
88	7	1.3	142	12	US-10-425-114-69680	

89	7	1.3	146	12	US-10-424-599-167688	Sequence 167688,	162	7	1.3	350	12	US-10-425-114-52194	Sequence 52194, A
90	7	1.3	156	10	US-09-893-519A-11	Sequence 11, Appl	163	7	1.3	350	14	US-10-309-851-8	Sequence 8, Appl
91	7	1.3	159	9	US-09-934-289A-20	Sequence 20, Appl	164	7	1.3	351	14	US-10-282-122A-76372	Sequence 76372, A
92	7	1.3	166	12	US-10-424-599-163595	Sequence 163595,	165	7	1.3	352	12	US-10-424-599-180936	Sequence 180936,
93	7	1.3	167	12	US-10-424-599-241739	Sequence 241739,	166	7	1.3	352	12	US-10-425-114-57918	Sequence 57918, A
94	7	1.3	171	12	US-10-424-599-262163	Sequence 262163,	167	7	1.3	352	12	US-10-425-114-46810	Sequence 46810, A
95	7	1.3	176	10	US-09-847-208-129	Sequence 129, App	168	7	1.3	355	12	US-10-425-114-64921	Sequence 64921, A
96	7	1.3	186	12	US-10-276-774-2467	Sequence 2467, Ap	169	7	1.3	356	12	US-10-425-114-47534	Sequence 47534, A
97	7	1.3	197	9	US-09-934-289A-18	Sequence 18, Appl	170	7	1.3	362	15	US-10-369-493-13566	Sequence 13566, A
98	7	1.3	206	12	US-10-425-114-45419	Sequence 45419, A	171	7	1.3	363	12	US-10-282-122A-54378	Sequence 54378, A
99	7	1.3	207	15	US-10-104-047-2607	Sequence 2607, Ap	172	7	1.3	365	12	US-10-424-599-168724	Sequence 168724, A
100	7	1.3	208	12	US-10-424-599-244841	Sequence 244841,	173	7	1.3	366	14	US-10-156-761-10482	Sequence 10482, A
101	7	1.3	215	10	US-09-866-480-339	Sequence 339, App	174	7	1.3	368	15	US-10-264-237-2400	Sequence 2400, Ap
102	7	1.3	215	10	US-09-892-877-359	Sequence 359, App	175	7	1.3	379	14	US-10-050-882-53	Sequence 53, Appl
103	7	1.3	215	10	US-09-948-783-163	Sequence 363, App	176	7	1.3	380	9	US-09-729-674-128	Sequence 128, App
104	7	1.3	215	12	US-10-425-114-43098	Sequence 43098, A	177	7	1.3	381	12	US-10-412-699B-606	Sequence 606, App
105	7	1.3	220	9	US-09-938-803-4	Sequence 4, Appl	178	7	1.3	381	14	US-10-295-403-112	Sequence 112, App
106	7	1.3	228	12	US-10-282-122A-48054	Sequence 48054, A	179	7	1.3	386	15	US-10-004-378A-4	Sequence 4, Appl
107	7	1.3	233	12	US-10-425-114-56502	Sequence 56502, A	180	7	1.3	386	15	US-10-369-493-7832	Sequence 7832, Ap
108	7	1.3	238	12	US-10-282-122A-48996	Sequence 48996, A	181	7	1.3	390	12	US-10-282-122A-65259	Sequence 65259, A
109	7	1.3	240	12	US-10-282-122A-43308	Sequence 43308, A	182	7	1.3	390	12	US-10-282-122A-65956	Sequence 65956, A
110	7	1.3	240	14	US-10-309-851-4	Sequence 4, Appl	183	7	1.3	401	12	US-10-425-114-37468	Sequence 37468, A
111	7	1.3	240	14	US-10-309-851-36	Sequence 36, Appl	184	7	1.3	403	15	US-10-369-493-10660	Sequence 10660, A
112	7	1.3	241	12	US-10-016-248-54	Sequence 54, Appl	185	7	1.3	406	12	US-10-335-977-7931	Sequence 7931, Ap
113	7	1.3	251	12	US-10-412-699B-536	Sequence 536, App	186	7	1.3	416	9	US-09-925-300-1395	Sequence 1395, Ap
114	7	1.3	251	14	US-10-295-403-142	Sequence 142, App	187	7	1.3	418	12	US-10-282-122A-49150	Sequence 49150, A
115	7	1.3	253	12	US-10-424-599-244838	Sequence 244838,	188	7	1.3	418	12	US-10-424-599-183808	Sequence 183808,
116	7	1.3	253	15	US-10-369-493-22886	Sequence 22886, A	189	7	1.3	420	12	US-10-424-599-285648	Sequence 285648,
117	7	1.3	258	12	US-10-016-248-53	Sequence 53, Appl	190	7	1.3	430	12	US-10-425-114-36713	Sequence 36713, A
118	7	1.3	263	12	US-10-425-114-47632	Sequence 47632, A	191	7	1.3	436	15	US-10-104-047-3432	Sequence 3432, Ap
119	7	1.3	264	12	US-10-282-122A-59293	Sequence 59293, A	192	7	1.3	437	12	US-10-425-114-55188	Sequence 55188, A
120	7	1.3	265	14	US-10-309-851-6	Sequence 6, Appl	193	7	1.3	446	9	US-09-824-735-2	Sequence 2, Appl
121	7	1.3	266	12	US-10-380-703-7	Sequence 7, Appl	194	7	1.3	446	12	US-10-617-038-8	Sequence 8, Appl
122	7	1.3	270	14	US-10-156-761-11866	Sequence 11866, A	195	7	1.3	451	15	US-10-004-378A-2	Sequence 2, Appl
123	7	1.3	271	12	US-10-424-599-199397	Sequence 199397,	196	7	1.3	453	12	US-10-282-122A-52293	Sequence 52293, A
124	7	1.3	274	12	US-10-424-599-224138	Sequence 224138,	197	7	1.3	453	12	US-09-978-295A-430	Sequence 430, App
125	7	1.3	275	12	US-10-282-122A-55720	Sequence 55720, A	198	7	1.3	455	9	US-09-978-697-430	Sequence 430, App
126	7	1.3	276	15	US-10-104-047-2279	Sequence 2279, Ap	199	7	1.3	455	9	US-09-978-192A-430	Sequence 430, App
127	7	1.3	277	12	US-10-282-122A-59764	Sequence 59764, A	200	7	1.3	455	9	US-09-965-529-14	Sequence 14, Appl
128	7	1.3	277	14	US-10-106-698-5823	Sequence 5823, Ap	201	7	1.3	455	9	US-09-999-832A-430	Sequence 430, App
129	7	1.3	280	12	US-10-282-122A-78227	Sequence 78227, A	202	7	1.3	455	10	US-09-978-183-430	Sequence 430, App
130	7	1.3	281	12	US-10-335-977-8435	Sequence 8435, Ap	203	7	1.3	455	10	US-09-978-608A-430	Sequence 430, App
131	7	1.3	282	12	US-10-425-114-46688	Sequence 46688, A	204	7	1.3	455	10	US-09-978-585A-430	Sequence 430, App
132	7	1.3	284	12	US-10-424-599-279326	Sequence 279326,	205	7	1.3	455	10	US-09-978-191A-430	Sequence 430, App
133	7	1.3	286	12	US-10-424-599-224134	Sequence 224134,	206	7	1.3	455	10	US-09-978-403A-430	Sequence 430, App
134	7	1.3	292	14	US-10-050-882-78	Sequence 78, Appl	207	7	1.3	455	10	US-09-978-564A-430	Sequence 430, App
135	7	1.3	293	14	US-10-017-161-1866	Sequence 1866, Ap	208	7	1.3	455	10	US-09-999-832A-430	Sequence 430, App
136	7	1.3	293	15	US-10-292-798-1522	Sequence 1522, Ap	209	7	1.3	455	10	US-09-981-918A-430	Sequence 430, App
137	7	1.3	296	12	US-10-182-936A-102	Sequence 102, App	210	7	1.3	455	10	US-09-978-757A-430	Sequence 430, App
138	7	1.3	296	15	US-10-374-979-102	Sequence 102, App	211	7	1.3	455	10	US-09-978-187B-430	Sequence 430, App
139	7	1.3	304	15	US-10-369-493-11313	Sequence 11313, A	212	7	1.3	455	10	US-09-978-643A-430	Sequence 430, App
140	7	1.3	306	15	US-09-815-242-11485	Sequence 11485, A	213	7	1.3	455	10	US-09-978-824-430	Sequence 430, App
141	7	1.3	309	9	US-09-729-674-70	Sequence 70, Appl	214	7	1.3	455	10	US-09-918-585A-430	Sequence 430, App
142	7	1.3	310	15	US-10-369-493-4700	Sequence 4700, Ap	215	7	1.3	455	10	US-09-978-423A-430	Sequence 430, App
143	7	1.3	310	15	US-10-104-047-3795	Sequence 3795, Ap	216	7	1.3	455	10	US-09-969-680A-14	Sequence 14, Appl
144	7	1.3	311	15	US-10-108-260A-2887	Sequence 2887, Ap	217	7	1.3	455	10	US-09-978-375A-430	Sequence 430, App
145	7	1.3	314	12	US-10-425-114-43684	Sequence 43684, A	218	7	1.3	455	10	US-09-978-298A-430	Sequence 430, App
146	7	1.3	316	12	US-10-424-599-256641	Sequence 256641,	219	7	1.3	455	10	US-09-978-298A-430	Sequence 430, App
147	7	1.3	317	12	US-10-282-122A-53118	Sequence 53118, A	220	7	1.3	455	10	US-09-978-188A-430	Sequence 430, App
148	7	1.3	323	12	US-10-425-114-42699	Sequence 42699, A	221	7	1.3	455	10	US-09-978-681A-430	Sequence 430, App
149	7	1.3	326	9	US-09-815-242-11485	Sequence 11485, A	222	7	1.3	455	10	US-09-978-681A-430	Sequence 430, App
150	7	1.3	326	12	US-09-882-227-502	Sequence 227-502	223	7	1.3	455	10	US-09-978-194A-430	Sequence 430, App
151	7	1.3	326	12	US-10-282-122A-58652	Sequence 58652, A	224	7	1.3	455	10	US-09-999-829A-430	Sequence 430, App
152	7	1.3	327	12	US-10-282-122A-76788	Sequence 76788, A	225	7	1.3	455	10	US-09-978-299A-430	Sequence 430, App
153	7	1.3	334	12	US-10-425-114-56973	Sequence 56973, A	226	7	1.3	455	10	US-09-978-544A-430	Sequence 430, App
154	7	1.3	337	12	US-10-424-599-271324	Sequence 271324,	227	7	1.3	455	10	US-09-978-665A-430	Sequence 430, App
155	7	1.3	337	14	US-10-097-111-295	Sequence 235, App	228	7	1.3	455	12	US-09-978-802A-430	Sequence 430, App
156	7	1.3	339	12	US-10-282-122A-51710	Sequence 51710, A	229	7	1.3	455	12	US-10-164-749A-430	Sequence 430, App
157	7	1.3	341	12	US-10-424-599-250813	Sequence 250813,	230	7	1.3	455	12	US-10-206-915-14	Sequence 14, Appl
158	7	1.3	341	12	US-10-424-599-250815	Sequence 250815,	231	7	1.3	455	12	US-10-199-670-14	Sequence 14, Appl
159	7	1.3	341	14	US-10-092-066B-1	Sequence 1, Appl	232	7	1.3	455	12	US-10-201-858-14	Sequence 14, Appl
160	7	1.3	343	12	US-10-424-599-266487	Sequence 266487,	233	7	1.3	455	12	US-09-999-831A-430	Sequence 430, App
161	7	1.3	350	12	US-10-424-599-180925	Sequence 180925,	234	7	1.3	455	12	US-10-205-890-14	Sequence 14, Appl

7	1.3	455	12	US-10-208-024-14	Sequence 14, Appl
235	7	455	12	US-10-201-853-14	Sequence 14, Appl
236	7	455	12	US-10-063-745-4	Sequence 4, Appl
237	7	455	12	US-10-013-917A-430	Sequence 430, App
238	7	455	12	US-10-063-512-4	Sequence 4, Appl
239	7	455	12	US-10-063-513-4	Sequence 4, Appl
240	7	455	12	US-10-063-515-4	Sequence 4, Appl
241	7	455	12	US-10-063-515-4	Sequence 4, Appl
242	7	455	12	US-10-063-549-4	Sequence 4, Appl
243	7	455	12	US-10-063-569-4	Sequence 4, Appl
244	7	455	12	US-10-063-551-4	Sequence 4, Appl
245	7	455	12	US-10-174-581-14	Sequence 14, Appl
246	7	455	12	US-10-176-483-14	Sequence 14, Appl
247	7	455	12	US-10-176-749-14	Sequence 14, Appl
248	7	455	12	US-10-176-914-14	Sequence 14, Appl
249	7	455	12	US-10-176-915-14	Sequence 14, Appl
250	7	455	12	US-09-999-834A-430	Sequence 430, App
251	7	455	12	US-10-063-555-4	Sequence 4, Appl
252	7	455	12	US-10-063-563-4	Sequence 4, Appl
253	7	455	12	US-10-063-594-4	Sequence 4, Appl
254	7	455	12	US-10-063-553-4	Sequence 4, Appl
255	7	455	12	US-10-063-554-4	Sequence 4, Appl
256	7	455	12	US-10-176-484-14	Sequence 14, Appl
257	7	455	12	US-10-180-550-14	Sequence 14, Appl
258	7	455	12	US-10-183-014-14	Sequence 14, Appl
259	7	455	12	US-10-187-738-14	Sequence 14, Appl
260	7	455	12	US-10-187-740-14	Sequence 14, Appl
261	7	455	12	US-10-187-883-14	Sequence 14, Appl
262	7	455	12	US-10-194-263-14	Sequence 14, Appl
263	7	455	12	US-10-194-460-14	Sequence 14, Appl
264	7	455	12	US-10-194-463-14	Sequence 14, Appl
265	7	455	12	US-10-194-484-14	Sequence 14, Appl
266	7	455	12	US-10-195-884-14	Sequence 14, Appl
267	7	455	12	US-10-195-896-14	Sequence 14, Appl
268	7	455	12	US-10-196-744-14	Sequence 14, Appl
269	7	455	12	US-10-196-755-14	Sequence 14, Appl
270	7	455	12	US-10-196-757-14	Sequence 14, Appl
271	7	455	12	US-10-197-704-14	Sequence 14, Appl
272	7	455	12	US-10-197-710-14	Sequence 14, Appl
273	7	455	12	US-10-198-758-14	Sequence 14, Appl
274	7	455	12	US-10-198-766-14	Sequence 14, Appl
275	7	455	12	US-10-199-304-14	Sequence 14, Appl
276	7	455	12	US-10-199-309-14	Sequence 14, Appl
277	7	455	12	US-10-199-313-14	Sequence 14, Appl
278	7	455	12	US-10-199-456-14	Sequence 14, Appl
279	7	455	12	US-10-201-329-14	Sequence 14, Appl
280	7	455	12	US-10-202-412-14	Sequence 14, Appl
281	7	455	12	US-10-206-919-14	Sequence 14, Appl
282	7	455	12	US-10-206-922-14	Sequence 14, Appl
283	7	455	12	US-10-206-924-14	Sequence 14, Appl
284	7	455	12	US-10-206-928-14	Sequence 14, Appl
285	7	455	12	US-10-207-914-14	Sequence 14, Appl
286	7	455	12	US-10-207-921-14	Sequence 14, Appl
287	7	455	12	US-10-207-922-14	Sequence 14, Appl
288	7	455	12	US-10-208-027-14	Sequence 14, Appl
289	7	455	12	US-10-182-521A-430	Sequence 430, App
290	7	455	12	US-10-174-570-14	Sequence 14, Appl
291	7	455	12	US-10-183-005-14	Sequence 14, Appl
292	7	455	12	US-10-145-016A-430	Sequence 430, App
293	7	455	12	US-10-145-088A-430	Sequence 430, App
294	7	455	12	US-10-145-092A-430	Sequence 430, App
295	7	455	12	US-10-145-129A-430	Sequence 430, App
296	7	455	12	US-10-185-038A-430	Sequence 430, App
297	7	455	12	US-10-165-353A-430	Sequence 430, App

ALIGNMENTS

```

1 Sequence 8, Application US/09962055
2 Patent No. US20020052033A1
3 GENERAL INFORMATION:
4 APPLICANT: Lees, Ann M.
5             Lees, Robert S.
6             Law, Simon W.
7             Arjona, Anibal A.
8 TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
9             BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
10            TREATING ATHEROSCLEROSIS
11
12 NUMBER OF SEQUENCES: 42
13 CORRESPONDENCE ADDRESS:
14 ADDRESSEE: Fish & Richardson P.C.
15 STREET: 225 Franklin Street
16 CITY: Boston
17 STATE: MA
18 COUNTRY: USA
19 ZIP: 02110-2804
20 COMPUTER READABLE FORM:
21 MEDIUM TYPE: Diskette
22 COMPUTER: IBM Compatible
23 OPERATING SYSTEM: DOS
24 SOFTWARE: FastSEQ for Windows Version 2.0
25 CURRENT APPLICATION DATA:
26 APPLICATION NUMBER: US/09/962,055
27 FILING DATE: 24-Sep-2001
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: 08/979,608
30 FILING DATE: 26-NOV-1997
31 APPLICATION NUMBER: US 60/031,930
32 FILING DATE: 27-NOV-1996
33 ATTORNEY/AGENT INFORMATION:
34 NAME: Myers, Louis
35 REGISTRATION NUMBER: 35,965
36 REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
37 TELECOMMUNICATION INFORMATION:
38 TELEPHONE: 617/542-5070
39 TELEFAX: 617/542-8906
40 INFORMATION FOR SEQ ID NO: 8:
41 SEQUENCE CHARACTERISTICS:
42     LENGTH: 530 amino acids
43     TYPE: amino acid
44     TOPOLOGY: linear
45 MOLECULE TYPE: protein
46 SEQUENCE DESCRIPTION: SEQ ID NO: 8:
47 US-09-962-055-8

```

Db 301 ELREEHIDKVFPHKDLQOOLVDAKLOQAQEMKAESEHOREKDFLLKEAVESQRMCELM 360
Qy 361 KOETHLKOOLALYTEKEFEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
Db 361 KOETHLKOOLALYTEKEFEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
Qy 421 ESSNKALLEMAEKTVRDKEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQGS 480
Db 421 ESSNKALLEMAEKTVRDKEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQGS 480
Qy 481 TDGSPRRPGRGAGAPSSPRVTEAPCPGAPSTASGOTGPQEPPTSARA 530
Db 481 TDGSPRRPGRGAGAPSSPRVTEAPCPGAPSTASGOTGPQEPPTSARA 530

RESULT 2
US-09-976-740-8
; Sequence 8, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-8

Query Match 100.0%; Score 530; DB 9; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSSPGQEPAGPGRGAGAPSSPRVTEAPCPGAPSTASGOTGPQEPPTSARA 60
Db 1 KSSPGQEPAGPGRGAGAPSSPRVTEAPCPGAPSTASGOTGPQEPPTSARA 60
Qy 61 RQLEDILSTYCVDNNGQGGEDGAGQEPAPEDAEKSRITYVARNGEPEPTPVYGEKPS 120
Db 61 RQLEDILSTYCVDNNGQGGEDGAGQEPAPEDAEKSRITYVARNGEPEPTPVYGEKPS 120
Qy 121 KGPNTTEIRQSDVEGDRDHRPQEKKAQGLGKEITLLMOTLNTLSTPEKLAALCKY 180
Db 121 KGPNTTEIRQSDVEGDRDHRPQEKKAQGLGKEITLLMOTLNTLSTPEKLAALCKY 180
Qy 181 AELLLEHNRNOKOMKLLQKQSQQLVQEKDHLRGHSHKAVLARSKLESICRELQHNRSK 240
Db 181 AELLLEHNRNOKOMKLLQKQSQQLVQEKDHLRGHSHKAVLARSKLESICRELQHNRSK 240
Qy 241 EGVQARAREEBEERKKEVTSHFQVTLNDIQLQWQHNRNNSKLQENMELARLKKLIQY 300
Db 241 EGVQARAREEBEERKKEVTSHFQVTLNDIQLQWQHNRNNSKLQENMELARLKKLIQY 300
Qy 301 ELREEHIDKVFPHKDLQOOLVDAKLOQAQEMKAESEHOREKDFLLKEAVESQRMCELM 360
Db 301 ELREEHIDKVFPHKDLQOOLVDAKLOQAQEMKAESEHOREKDFLLKEAVESQRMCELM 360

Qy 361 KOETHLKOOLALYTEKEFEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
Db 361 KOETHLKOOLALYTEKEFEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
Qy 421 ESSNKALLEMAEKTVRDKEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQGS 480
Db 421 ESSNKALLEMAEKTVRDKEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQGS 480
Qy 481 TDGSPRRPGRGAGAPSSPRVTEAPCPGAPSTASGOTGPQEPPTSARA 530
Db 481 TDGSPRRPGRGAGAPSSPRVTEAPCPGAPSTASGOTGPQEPPTSARA 530

RESULT 3
US-10-671-242-8
; Sequence 8, Application US/10671242
; Publication No. US20040040049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; PRIOR FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-8

Query Match 100.0%; Score 530; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSSPGQEPAGPGRGAGAPSSPRVTEAPCPGAPSTASGOTGPQEPPTSARA 60
Db 1 KSSPGQEPAGPGRGAGAPSSPRVTEAPCPGAPSTASGOTGPQEPPTSARA 60
Qy 61 RQLEDILSTYCVDNNGQGGEDGAGQEPAPEDAEKSRITYVARNGEPEPTPVYGEKPS 120
Db 61 RQLEDILSTYCVDNNGQGGEDGAGQEPAPEDAEKSRITYVARNGEPEPTPVYGEKPS 120
Qy 121 KGPNTTEIRQSDVEGDRDHRPQEKKAQGLGKEITLLMOTLNTLSTPEKLAALCKY 180
Db 121 KGPNTTEIRQSDVEGDRDHRPQEKKAQGLGKEITLLMOTLNTLSTPEKLAALCKY 180
Qy 181 AELLLEHNRNOKOMKLLQKQSQQLVQEKDHLRGHSHKAVLARSKLESICRELQHNRSK 240
Db 181 AELLLEHNRNOKOMKLLQKQSQQLVQEKDHLRGHSHKAVLARSKLESICRELQHNRSK 240
Qy 241 EGVQARAREEBEERKKEVTSHFQVTLNDIQLQWQHNRNNSKLQENMELARLKKLIQY 300
Db 241 EGVQARAREEBEERKKEVTSHFQVTLNDIQLQWQHNRNNSKLQENMELARLKKLIQY 300
Qy 301 ELREEHIDKVFPHKDLQOOLVDAKLOQAQEMKAESEHOREKDFLLKEAVESQRMCELM 360
Db 301 ELREEHIDKVFPHKDLQOOLVDAKLOQAQEMKAESEHOREKDFLLKEAVESQRMCELM 360

Db 301 ELREEHIDKVFHKDIQQQLVDAKIQQAQEMLKEABERHQRKDFLLKEAVESQRMCELM 360
Qy 361 KQETHLKOALALYTFKPEPONTLSKSSSEVFTTFKQEMEMTKTKIKLEKETMYRSRW 420
Db 361 KQETHLKOALALYTFKPEPONTLSKSSSEVFTTFKQEMEMTKTKIKLEKETMYRSRW 420
Qy 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSSL 480
Db 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSSL 480
Qy 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQPTTSARA 530
Db 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQPTTSARA 530

RESULT 4
US-10-023-529-8
; Sequence 8, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-529-8

Query Match 100.0%; Score 530; DB 13; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEGAQARTAQSGALRDVSELS 60
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEGAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVDNNGGPGEDGAGPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Db 61 RQLEDILSTYCVDNNGGPGEDGAGPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Qy 121 KGPNTPEIRQSDVGDGRDHRHPQEKKAAGLGEKTEILLMOTLNTLSTPEEKLAALCKKY 180
Db 121 KGPNTPEIRQSDVGDGRDHRHPQEKKAAGLGEKTEILLMOTLNTLSTPEEKLAALCKKY 180
Qy 181 AELLBEHRNSOKMKLLQKKOSQLVQEKDHLRGHSAVLARSKLESCRELQHRNLSK 240
Db 181 AELLBEHRNSOKMKLLQKKOSQLVQEKDHLRGHSAVLARSKLESCRELQHRNLSK 240
Qy 241 EEGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 300
Db 241 EEGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 300
Qy 301 ELREEHIDKVFHKDIQQQLVDAKIQQAQEMLKEABERHQRKDFLLKEAVESQRMCELM 360

Db 301 ELREEHIDKVFHKDIQQQLVDAKIQQAQEMLKEABERHQRKDFLLKEAVESQRMCELM 360
Qy 361 KQETHLKOALALYTFKPEPONTLSKSSSEVFTTFKQEMEMTKTKIKLEKETMYRSRW 420
Db 361 KQETHLKOALALYTFKPEPONTLSKSSSEVFTTFKQEMEMTKTKIKLEKETMYRSRW 420
Qy 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSSL 480
Db 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSSL 480
Qy 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQPTTSARA 530
Db 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQPTTSARA 530

RESULT 5
US-10-023-523-8
; Sequence 8, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-523-8

Query Match 100.0%; Score 530; DB 13; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEGAQARTAQSGALRDVSELS 60
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEGAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVDNNGGPGEDGAGPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Db 61 RQLEDILSTYCVDNNGGPGEDGAGPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Qy 121 KGPNTPEIRQSDVGDGRDHRHPQEKKAAGLGEKTEILLMOTLNTLSTPEEKLAALCKKY 180
Db 121 KGPNTPEIRQSDVGDGRDHRHPQEKKAAGLGEKTEILLMOTLNTLSTPEEKLAALCKKY 180
Qy 181 AELLBEHRNSOKMKLLQKKOSQLVQEKDHLRGHSAVLARSKLESCRELQHRNLSK 240
Db 181 AELLBEHRNSOKMKLLQKKOSQLVQEKDHLRGHSAVLARSKLESCRELQHRNLSK 240
Qy 241 EEGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 300
Db 241 EEGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 300

Qy 301 ELREHIDKVPKHKDLQOOLVDKALQOQAEMLKEABERHOREKDFLLKEAVESORMCELM 360
Db 301 ELREHIDKVPKHKDLQOOLVDKALQOQAEMLKEABERHOREKDFLLKEAVESORMCELM 360
Qy 361 KQETHLKQOALALYTEKEFEFQNTLSKSSEVFTTFKQEMERMTKKIKKLEKETTMYRSRW 420
Db 361 KQETHLKQOALALYTEKEFEFQNTLSKSSEVFTTFKQEMERMTKKIKKLEKETTMYRSRW 420
Qy 421 ESSNKALLEMAEBKTVROKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
Db 421 ESSNKALLEMAEBKTVROKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
Qy 481 TDSGPERPEGPGAGAPSSPRVTEAPCYPGAPSTEASGOTGPQPTTSARA 530
Db 481 TDSGPERPEGPGAGAPSSPRVTEAPCYPGAPSTEASGOTGPQPTTSARA 530

RESULT 6

US-10-616-187-8
; Sequence 8, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US/09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-8

Query Match 100.0%; Score 530; DB 15; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKEGAQARTAQSGALRDVSELS 60
Db 1 KSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKEGAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVNNQGGGEGDGAQGEPAEPDAEKSRITYVARNGEPTPTVYVGEKPS 120
Db 61 RQLEDILSTYCVNNQGGGEGDGAQGEPAEPDAEKSRITYVARNGEPTPTVYVGEKPS 120
Qy 121 KGPDPTEEIRQSDVEGDRDHRPQEKKAAGLGEITLLMOTLNTLSTPEKLAALCKKY 180
Db 121 KGPDPTEEIRQSDVEGDRDHRPQEKKAAGLGEITLLMOTLNTLSTPEKLAALCKKY 180
Qy 181 AELLBEHRNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSKLESICRELQHRNSLK 240
Db 181 AELLBEHRNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSKLESICRELQHRNSLK 240
Qy 241 EBGVQARAEERKEKVTSHFQVTLNDIQLQVQEHNRNSKLRQENMELAEKLIQY 300
Db 241 EBGVQARAEERKEKVTSHFQVTLNDIQLQVQEHNRNSKLRQENMELAEKLIQY 300

Qy 301 ELREHIDKVPKHKDLQOOLVDKALQOQAEMLKEABERHOREKDFLLKEAVESORMCELM 360
Db 301 ELREHIDKVPKHKDLQOOLVDKALQOQAEMLKEABERHOREKDFLLKEAVESORMCELM 360
Qy 361 KQETHLKQOALALYTEKEFEFQNTLSKSSEVFTTFKQEMERMTKKIKKLEKETTMYRSRW 420
Db 361 KQETHLKQOALALYTEKEFEFQNTLSKSSEVFTTFKQEMERMTKKIKKLEKETTMYRSRW 420
Qy 421 ESSNKALLEMAEBKTVROKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
Db 421 ESSNKALLEMAEBKTVROKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
Qy 481 TDSGPERPEGPGAGAPSSPRVTEAPCYPGAPSTEASGOTGPQPTTSARA 530
Db 481 TDSGPERPEGPGAGAPSSPRVTEAPCYPGAPSTEASGOTGPQPTTSARA 530

RESULT 7

US-09-976-740-44
; Sequence 44, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-44

Query Match 80.9%; Score 429; DB 9; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 KSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKEGAQARTAQSGALRDVSELS 60
Db 17 KSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKEGAQARTAQSGALRDVSELS 76
Qy 61 RQLEDILSTYCVNNQGGGEGDGAQGEPAEPDAEKSRITYVARNGEPTPTVYVGEKPS 120
Db 77 RQLEDILSTYCVNNQGGGEGDGAQGEPAEPDAEKSRITYVARNGEPTPTVYVGEKPS 136
Qy 121 KGPDPTEEIRQSDVEGDRDHRPQEKKAAGLGEITLLMOTLNTLSTPEKLAALCKKY 180
Db 137 KGPDPTEEIRQSDVEGDRDHRPQEKKAAGLGEITLLMOTLNTLSTPEKLAALCKKY 196
Qy 181 AELLBEHRNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSKLESICRELQHRNSLK 240
Db 197 AELLBEHRNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSKLESICRELQHRNSLK 256
Qy 241 EBGVQARAEERKEKVTSHFQVTLNDIQLQVQEHNRNSKLRQENMELAEKLIQY 300
Db 257 EBGVQARAEERKEKVTSHFQVTLNDIQLQVQEHNRNSKLRQENMELAEKLIQY 316

```

QY 301 ELREHIDKVFHKDLOQOLVDKLOQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
DB 317 ELREHIDKVFHKDLOQOLVDKLOQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 376
QY 361 KQOETHLKQOLALYTEKPEEFONTLSKSSSEVFTTFKQEMERKTKIKKLEKETTYMRSRW 420
DB 377 KQOETHLKQOLALYTEKPEEFONTLSKSSSEVFTTFKQEMERKTKIKKLEKETTYMRSRW 436
QY 421 ESSNKALLEMABEKTVRDKELEGLQVKIORLEKLCALQTERNDLNKRVODLSAGQGSLS 480
DB 437 ESSNKALLEMABEKTVRDKELEGLQVKIORLEKLCALQTERNDLNKRVODLSAGQGSLS 496
QY 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPPTSARA 530
DB 497 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPPTSARA 546

```

```

RESULT 8
US-10-671-242-44
; Sequence 44, Application US/10671242
; Publication No. US2004040049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; CURRENT FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-44

```

```

Query Match 80.9%; Score 429; DB 12; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEPGAQARTAQSGALRDVSEELS 60
DB 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEPGAQARTAQSGALRDVSEELS 76
QY 61 ROLEDILSTYCVNNQGGPGEAGQEPAPEDAEKSRITYVARNGEPEPTPVYVGEKPS 120
DB 77 ROLEDILSTYCVNNQGGPGEAGQEPAPEDAEKSRITYVARNGEPEPTPVYVGEKPS 136
QY 121 KGDPTNTEIRQSDVEGDRHRRPQKKKAGLGEITLLMOTLNTLSTPEEKLAALCKKY 180
DB 137 KGDPTNTEIRQSDVEGDRHRRPQKKKAGLGEITLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 AELLLEHNSQKMKLQKQSOVLQVQKHRLRGEHSKAVLARSLKSLCRLQHRNRSLK 240
DB 197 AELLLEHNSQKMKLQKQSOVLQVQKHRLRGEHSKAVLARSLKSLCRLQHRNRSLK 256
QY 241 EGVQVAREEERKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 300
DB 257 EGVQVAREEERKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 316

```

```

QY 301 ELREHIDKVFHKDLOQOLVDKLOQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
DB 317 ELREHIDKVFHKDLOQOLVDKLOQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 376
QY 361 KQOETHLKQOLALYTEKPEEFONTLSKSSSEVFTTFKQEMERKTKIKKLEKETTYMRSRW 420
DB 377 KQOETHLKQOLALYTEKPEEFONTLSKSSSEVFTTFKQEMERKTKIKKLEKETTYMRSRW 436
QY 421 ESSNKALLEMABEKTVRDKELEGLQVKIORLEKLCALQTERNDLNKRVODLSAGQGSLS 480
DB 437 ESSNKALLEMABEKTVRDKELEGLQVKIORLEKLCALQTERNDLNKRVODLSAGQGSLS 496
QY 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPPTSARA 530
DB 497 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPPTSARA 546

```

```

RESULT 9
US-10-023-529-44
; Sequence 44, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-529-44

```

```

Query Match 80.9%; Score 429; DB 13; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEPGAQARTAQSGALRDVSEELS 60
DB 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEPGAQARTAQSGALRDVSEELS 76
QY 61 ROLEDILSTYCVNNQGGPGEAGQEPAPEDAEKSRITYVARNGEPEPTPVYVGEKPS 120
DB 77 ROLEDILSTYCVNNQGGPGEAGQEPAPEDAEKSRITYVARNGEPEPTPVYVGEKPS 136
QY 121 KGDPTNTEIRQSDVEGDRHRRPQKKKAGLGEITLLMOTLNTLSTPEEKLAALCKKY 180
DB 137 KGDPTNTEIRQSDVEGDRHRRPQKKKAGLGEITLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 AELLLEHNSQKMKLQKQSOVLQVQKHRLRGEHSKAVLARSLKSLCRLQHRNRSLK 240
DB 197 AELLLEHNSQKMKLQKQSOVLQVQKHRLRGEHSKAVLARSLKSLCRLQHRNRSLK 256
QY 241 EGVQVAREEERKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 300

```

Db 257 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQENMELAEERLKKLIEQY 316
QY 301 ELREEHIDKVFKHDLQOOLVDKILQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM 360
Db 317 ELREEHIDKVFKHDLQOOLVDKILQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM 376
QY 361 KQETHLKOQALALYTKFPEFQNTLSKSEVFTTFKQEMEMOTKKIKLEKETTTMYRSRW 420
Db 377 KQETHLKOQALALYTKFPEFQNTLSKSEVFTTFKQEMEMOTKKIKLEKETTTMYRSRW 436
QY 421 ESSNKALLEMAEERTVDRKEGLQVKTORLEKLCRALQTERNDLNKRVODLSAGGGSLS 480
Db 437 ESSNKALLEMAEERTVDRKEGLQVKTORLEKLCRALQTERNDLNKRVODLSAGGGSLS 496
QY 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPSTARA 530
Db 497 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPSTARA 546

RESULT 10
US-10-023-523-44
; Sequence 44, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1998-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-523-44

Query Match 80.9%; Score 429; DB 13; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKPGEQAQARTAGSALRDVSELS 60
Db 17 KSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKPGEQAQARTAGSALRDVSELS 76
QY 61 RQLEDILSTVCVDNNGGCGEDGAGCEPAEPDAEKSRITYVARNGEPEPTPVVYGEKEPS 120
Db 77 RQLEDILSTVCVDNNGGCGEDGAGCEPAEPDAEKSRITYVARNGEPEPTPVVYGEKEPS 136
QY 121 KGDPTNTEIRQSDVEGDRDHRHPQEKKAAGLGEITLLMQTLNTLSTPEKLAALCKKY 180
Db 137 KGDPTNTEIRQSDVEGDRDHRHPQEKKAAGLGEITLLMQTLNTLSTPEKLAALCKKY 196
QY 181 AELLEHRNSOKMKLLQKKOSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNSLK 240
Db 197 AELLEHRNSOKMKLLQKKOSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNSLK 256
QY 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQENMELAEERLKKLIEQY 300

Db 257 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQENMELAEERLKKLIEQY 316
QY 301 ELREEHIDKVFKHDLQOOLVDKILQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM 360
Db 317 ELREEHIDKVFKHDLQOOLVDKILQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM 376
QY 361 KQETHLKOQALALYTKFPEFQNTLSKSEVFTTFKQEMEMOTKKIKLEKETTTMYRSRW 420
Db 377 KQETHLKOQALALYTKFPEFQNTLSKSEVFTTFKQEMEMOTKKIKLEKETTTMYRSRW 436
QY 421 ESSNKALLEMAEERTVDRKEGLQVKTORLEKLCRALQTERNDLNKRVODLSAGGGSLS 480
Db 437 ESSNKALLEMAEERTVDRKEGLQVKTORLEKLCRALQTERNDLNKRVODLSAGGGSLS 496
QY 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPSTARA 530
Db 497 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPSTARA 546

RESULT 11
US-10-616-187-44
; Sequence 44, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1998-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-44

Query Match 80.9%; Score 429; DB 15; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKPGEQAQARTAGSALRDVSELS 60
Db 17 KSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKPGEQAQARTAGSALRDVSELS 76
QY 61 RQLEDILSTVCVDNNGGCGEDGAGCEPAEPDAEKSRITYVARNGEPEPTPVVYGEKEPS 120
Db 77 RQLEDILSTVCVDNNGGCGEDGAGCEPAEPDAEKSRITYVARNGEPEPTPVVYGEKEPS 136
QY 121 KGDPTNTEIRQSDVEGDRDHRHPQEKKAAGLGEITLLMQTLNTLSTPEKLAALCKKY 180
Db 137 KGDPTNTEIRQSDVEGDRDHRHPQEKKAAGLGEITLLMQTLNTLSTPEKLAALCKKY 196
QY 181 AELLEHRNSOKMKLLQKKOSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNSLK 240
Db 197 AELLEHRNSOKMKLLQKKOSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNSLK 256

241 EGVORAREBEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 300
257 EGVORAREBEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 316
301 ELREEHIDKVFKHDIQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
317 ELREEHIDKVFKHDIQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376
361 KQOETHLKOALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRW 420
377 KQOETHLKOALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRW 436
421 ESSNKALLEAEBEKTVDKLEGLQVKIQLEKLCRALQTERNDLNKRVQDLGAGGQSL 480
437 ESSNKALLEAEBEKTVDKLEGLQVKIQLEKLCRALQTERNDLNKRVQDLGAGGQSL 496
481 TDSGPERRPEGQAGAPSSPRVTAPCPYGPAPSTASGOTGPQPTPSARA 530
497 TDSGPERRPEGQAGAPSSPRVTAPCPYGPAPSTASGOTGPQPTPSARA 546

RESULT 12
US-10-276-774-2134
; Sequence 2134, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y. Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2134
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(510)
; OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-2134

Query Match 67.7%; Score 359; DB 12; Length 510;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERFSQAAPAVEAEGPGSSQAAPRKEGAQARTAGSALRDVSELS 60
DB 28 KSSPGQPEAGPEGAQERFSQAAPAVEAEGPGSSQAAPRKEGAQARTAGSALRDVSELS 87
QY 61 RQLEDILSTYCVDNQGGPGEAGCAEPAEDAEKSTYVARNGEPEPTPVNGEKEPS 120
DB 88 RQLEDILSTYCVDNQGGPGEAGCAEPAEDAEKSTYVARNGEPEPTPVNGEKEPS 147
QY 121 KGDPNTEIRQSDVEGDRHRRPQKQKAKGLGKEITLLMOTLNTLSTPEEKLALCKKY 180
DB 148 KGDPNTEIRQSDVEGDRHRRPQKQKAKGLGKEITLLMOTLNTLSTPEEKLALCKKY 207
QY 181 AELLLEHNRNSQOMKLLQKQSQVQKDHLEGEHNSKAVLARSLSLRELOHNRSLK 240
DB 208 AELLLEHNRNSQOMKLLQKQSQVQKDHLEGEHNSKAVLARSLSLRELOHNRSLK 267
QY 241 EGVORAREBEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 300
DB 268 EGVORAREBEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 327
QY 301 ELREEHIDKVFKHDIQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360

DB 328 ELREEHIDKVFKHDIQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 387
QY 361 KQOETHLKOALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRW 420
DB 388 KQOETHLKOALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRW 447
QY 421 ESSNKALLEAEBEKTVDKLEGLQVKIQLEKLCRALQTERNDLNKRVQDLGAGGQSL 480
DB 448 ESSNKALLEAEBEKTVDKLEGLQVKIQLEKLCRALQTERNDLNKRVQDLGAGGQSL 496

RESULT 13
US-09-962-055-5
; Sequence 5, Application US/09962055
; Patent No. US20020052033A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M. S.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Aribal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,055
; FILING DATE: 24-Sep-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-962-055-5

Query Match 32.5%; Score 172; DB 9; Length 557;
Best Local Similarity 100.0%; Pred. No. 3,1e-155;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 TLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQYELREEHIDKVFKHDIQQLVDA 323
DB 280 TLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQYELREEHIDKVFKHDIQQLVDA 339
QY 324 KQOETHLKOALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRW 383
DB 340 KQOETHLKOALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRW 399

```
QY 384 T L S K S S E V T T F K Q E M E K T K I K L E K E T T M Y R S R W S S N K A L L E M A E E K T 435
Db 400 T L S K S S E V T T F K Q E M E K T K I K L E K E T T M Y R S R W S S N K A L L E M A E E K T 451

RESULT 14
US-09-976-740-5
; Sequence 5, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-09-976-740-5

Query Match 32.5%; Score 172; DB 9; Length 557;
Best Local Similarity 100.0%; Pred. No. 3.1e-155; Indels 0; Gaps 0;
Matches 172; Conservative 0; Mismatches 0;

QY 264 T L N D I Q L Q M E Q H N R N S K L R Q E N M E L A E R L K K L I E Q Y E L R E H I D K V F K H D L Q Q L V D A 323
Db 280 T L N D I Q L Q M E Q H N R N S K L R Q E N M E L A E R L K K L I E Q Y E L R E H I D K V F K H D L Q Q L V D A 339
QY 324 K L Q Q A Q E M L K E A E R H Q R E K D F L L K E A V E S Q R M C E L M K Q Q E T H L K Q Q L A L Y T E K F E E F Q N 383
Db 340 K L Q Q A Q E M L K E A E R H Q R E K D F L L K E A V E S Q R M C E L M K Q Q E T H L K Q Q L A L Y T E K F E E F Q N 399
QY 384 T L S K S S E V T T F K Q E M E K T K I K L E K E T T M Y R S R W S S N K A L L E M A E E K T 435
Db 400 T L S K S S E V T T F K Q E M E K T K I K L E K E T T M Y R S R W S S N K A L L E M A E E K T 451

RESULT 15
US-10-671-242-5
; Sequence 5, Application US/10671242
; Publication No. US20040040049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; CURRENT FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
```

```
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-671-242-5

Query Match 32.5%; Score 172; DB 12; Length 557;
Best Local Similarity 100.0%; Pred. No. 3.1e-155; Indels 0; Gaps 0;
Matches 172; Conservative 0; Mismatches 0;

QY 264 T L N D I Q L Q M E Q H N R N S K L R Q E N M E L A E R L K K L I E Q Y E L R E H I D K V F K H D L Q Q L V D A 323
Db 280 T L N D I Q L Q M E Q H N R N S K L R Q E N M E L A E R L K K L I E Q Y E L R E H I D K V F K H D L Q Q L V D A 339
QY 324 K L Q Q A Q E M L K E A E R H Q R E K D F L L K E A V E S Q R M C E L M K Q Q E T H L K Q Q L A L Y T E K F E E F Q N 383
Db 340 K L Q Q A Q E M L K E A E R H Q R E K D F L L K E A V E S Q R M C E L M K Q Q E T H L K Q Q L A L Y T E K F E E F Q N 399
QY 384 T L S K S S E V T T F K Q E M E K T K I K L E K E T T M Y R S R W S S N K A L L E M A E E K T 435
Db 400 T L S K S S E V T T F K Q E M E K T K I K L E K E T T M Y R S R W S S N K A L L E M A E E K T 451

RESULT 16
US-10-023-529-5
; Sequence 5, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-023-529-5

Query Match 32.5%; Score 172; DB 13; Length 557;
Best Local Similarity 100.0%; Pred. No. 3.1e-155; Indels 0; Gaps 0;
Matches 172; Conservative 0; Mismatches 0;

QY 264 T L N D I Q L Q M E Q H N R N S K L R Q E N M E L A E R L K K L I E Q Y E L R E H I D K V F K H D L Q Q L V D A 323
Db 280 T L N D I Q L Q M E Q H N R N S K L R Q E N M E L A E R L K K L I E Q Y E L R E H I D K V F K H D L Q Q L V D A 339
QY 324 K L Q Q A Q E M L K E A E R H Q R E K D F L L K E A V E S Q R M C E L M K Q Q E T H L K Q Q L A L Y T E K F E E F Q N 383
```



```
Db 340 KLOQAQEMLKAEERHOREKDFLLKEAVESQRCMELMKQETHLKQQLALYTEKPEEFQ 399
OY 384 TLSKSEVFTTFQEMEKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 435
Db 400 TLSKSEVFTTFQEMEKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 451

RESULT 17
US-10-023-523-5
; Sequence 5, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: ATHEROSCLEROSIS
; CURRENT APPLICATION NUMBER: US/10/023,523
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-023-523-5

Query Match 32.5%; Score 172; DB 13; Length 557;
Best Local Similarity 100.0%; Pred. No. 3.1e-155; Indels 0; Gaps 0;
Matches 172; Conservative 0; Mismatches 0;

OY 264 TLNDIQLQMEQHNRNSKLRQENMELARLKKLIQYELREEHIDKVPFKHDKLQQLVDA 323
Db 280 TLNDIQLQMEQHNRNSKLRQENMELARLKKLIQYELREEHIDKVPFKHDKLQQLVDA 339
OY 324 KLOQAQEMLKAEERHOREKDFLLKEAVESQRCMELMKQETHLKQQLALYTEKPEEFQ 383
Db 340 KLOQAQEMLKAEERHOREKDFLLKEAVESQRCMELMKQETHLKQQLALYTEKPEEFQ 399
OY 384 TLSKSEVFTTFQEMEKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 435
Db 400 TLSKSEVFTTFQEMEKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 451

RESULT 19
US-10-023-523-5
; Sequence 5, Application US/10616187
; Publication No. US20040013669A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: ATHEROSCLEROSIS
; CURRENT APPLICATION NUMBER: US/10/616,187
; PRIOR FILING DATE: 2003-07-09
```

```
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-616-187-5

Query Match 32.5%; Score 172; DB 15; Length 557;
Best Local Similarity 100.0%; Pred. No. 3.1e-155; Indels 0; Gaps 0;
Matches 172; Conservative 0; Mismatches 0;

OY 264 TLNDIQLQMEQHNRNSKLRQENMELARLKKLIQYELREEHIDKVPFKHDKLQQLVDA 323
Db 280 TLNDIQLQMEQHNRNSKLRQENMELARLKKLIQYELREEHIDKVPFKHDKLQQLVDA 339
OY 324 KLOQAQEMLKAEERHOREKDFLLKEAVESQRCMELMKQETHLKQQLALYTEKPEEFQ 383
Db 340 KLOQAQEMLKAEERHOREKDFLLKEAVESQRCMELMKQETHLKQQLALYTEKPEEFQ 399
OY 384 TLSKSEVFTTFQEMEKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 435
Db 400 TLSKSEVFTTFQEMEKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 451

RESULT 19
US-10-264-049-3413
; Sequence 3413, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133PI
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 3413
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (3)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (75)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-3413
```


Query Match 3.2%; Score 17; DB 15; Length 85;
Best Local Similarity 100.0%; Pred. No. 6.3e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 397 QEMERWTKKIKKLEKET 413
|||||
Db 52 QEMERWTKKIKKLEKET 68

RESULT 20
US-10-424-599-207434
; Sequence 207434, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 207434
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(436)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_29338C.1.pep
US-10-424-599-207434

Query Match 2.1%; Score 11; DB 12; Length 436;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 KLESCLRELQR 234
|||||
Db 194 KLESCLRELQR 204

RESULT 21
US-09-764-869-1082
; Sequence 1082, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1082
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (45)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (90)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (97)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-869-1082

Query Match 1.7%; Score 9; DB 9; Length 99;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GEKEPSKGD 123
|||||
Db 49 GEKEPSKGD 57

RESULT 22
US-10-091-504-1082
; Sequence 1082, Application US/10091504
; Publication No. US2003005908A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C1
; CURRENT APPLICATION NUMBER: US/10/091,504
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2442
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1082
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc feature
; LOCATION: (90)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc feature
; LOCATION: (97)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-091-504-1082

Query Match 1.7%; Score 9; DB 14; Length 99;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GEKEPSKGD 123
|||||
Db 49 GEKEPSKGD 57

RESULT 23
US-10-227-577-1082
; Sequence 1082, Application US/10227577
; Publication No. US20040005575A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C2
; CURRENT APPLICATION NUMBER: US/10/227,577
; CURRENT FILING DATE: 2002-08-26
; Prior Application Number: 10/091,504
; Prior Filing Date: 2002-03-07
; Prior Application Number: 09/764,869
; Prior Filing Date: 2001-01-17
; Prior Application Number: 60/179,065
; Prior Filing Date: 2000-01-31
; Prior Application Number: 60/180,628
; Prior Filing Date: 2000-02-04
; Prior Application Number: 60/214,886
; Prior Filing Date: 2000-06-28
; Prior Application Number: 60/217,487
; Prior Filing Date: 2000-07-11
; Prior Application Number: 60/225,758
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/220,963

PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/217,496
PRIOR FILING DATE: 2000-07-21
PRIOR APPLICATION NUMBER: 60/225,447
PRIOR FILING DATE: 2000-08-14
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1082
LENGTH: 99
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (45)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: misc feature
LOCATION: (90)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: misc feature
LOCATION: (97)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-227-577-1082

Query Match 1.7%; Score 9; DB 15; Length 99;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GEKEPSKGD 123
Db 49 GEKEPSKGD 57

RESULT 24
US-09-801-368-112
Sequence 112, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Cali, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801.368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 112
LENGTH: 594
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-112

Query Match 1.7%; Score 9; DB 9; Length 594;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 NTLSTPEEK 172
Db 466 NTLSTPEEK 474

RESULT 25
US-10-369-493-1762
Sequence 1762, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yougwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 1762
LENGTH: 594
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1762

Query Match 1.7%; Score 9; DB 15; Length 594;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 NTLSTPEEK 172
Db 466 NTLSTPEEK 474

RESULT 26
US-10-282-122A-72115
Sequence 72115, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangeu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA-034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27

```
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72115
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-282-122A-72115

Query Match      1.7%; Score 9; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 BELSROLED 65
Db 629 BELSROLED 637

RESULT 27
US-10-443-622-130
; Sequence 130, Application US/10443622
; Publication No. US20040024192A1
; GENERAL INFORMATION:
; APPLICANT: Carter et al.
; TITLE OF INVENTION: 19 Human Secreted Proteins
; FILE REFERENCE: PZ009P1
; CURRENT APPLICATION NUMBER: US/10/443,622
; CURRENT FILING DATE: 2003-05-23
; PRIOR APPLICATION NUMBER: PCT/US98/13608
; PRIOR FILING DATE: 1998-06-30
; PRIOR FILING DATE: 1997-07-01
; PRIOR FILING DATE: 1997-07-01
; PRIOR FILING DATE: 1997-09-12
; PRIOR FILING DATE: 1997-09-12
; PRIOR FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 130
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-443-622-130

Query Match      1.5%; Score 8; DB 16; Length 25;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 468 RVQDLSAG 475
Db 5 RVQDLSAG 12

RESULT 28
US-10-425-114-48446
; Sequence 48446, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
```

```
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48446
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700041743_FLI.pap
US-10-425-114-48446

Query Match      1.5%; Score 8; DB 12; Length 106;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GPEGAQER 17
Db 27 GPEGAQER 34

RESULT 29
US-10-424-599-272074
; Sequence 272074, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 272074
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_87701C.1.pap
US-10-424-599-272074

Query Match      1.5%; Score 8; DB 12; Length 121;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 PGSSQAPR 37
Db 70 PGSSQAPR 77

RESULT 30
US-10-443-622-122
; Sequence 122, Application US/10443622
; Publication No. US20040024192A1
; GENERAL INFORMATION:
; APPLICANT: Carter et al.
; TITLE OF INVENTION: 19 Human Secreted Proteins
; FILE REFERENCE: PZ009P1
; CURRENT APPLICATION NUMBER: US/10/443,622
; CURRENT FILING DATE: 2003-05-23
; PRIOR APPLICATION NUMBER: PCT/US98/13608
; PRIOR FILING DATE: 1998-06-30
; PRIOR FILING DATE: 1997-07-01
; PRIOR FILING DATE: 1997-07-01
; PRIOR FILING DATE: 1997-07-01
; PRIOR FILING DATE: 1997-09-12
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: 60/058,598
```

```
; PRIOR FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 122
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-443-622-122

Query Match      1.5%; Score 8; DB 16; Length 156;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      468 RVQDLSAG 475
Db      87 RVQDLSAG 94

RESULT 31
US-10-080-170-316
; Sequence 316, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TREATMENT OF MYCOBACTERIOSES
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495-0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 316
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-080-170-316

Query Match      1.5%; Score 8; DB 14; Length 229;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      18 PSQAAPAV 25
Db      200 PSQAAPAV 207

RESULT 32
US-10-156-761-12433
; Sequence 12433, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12433
; LENGTH: 413
; TYPE: PRT

; ORGANISM: Streptomyces avermitilis
US-10-156-761-12433

Query Match      1.5%; Score 8; DB 14; Length 413;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      49 SCALRDVS 56
Db      297 SCALRDVS 304

RESULT 33
US-10-369-493-21212
; Sequence 21212, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21212
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-369-493-21212

Query Match      1.5%; Score 8; DB 15; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      149 AKGLGKEI 156
Db      397 AKGLGKEI 404

RESULT 34
US-10-335-977-7865
; Sequence 7865, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
```

ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 7865:
SEQUENCE CHARACTERISTICS:
LENGTH: 450 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...450
SEQUENCE DESCRIPTION: SEQ ID NO: 7865:
US-10-335-977-7865
Query Match 1.5%; Score 8; DB 12; Length 450;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 63 LEDILSTY 70
Db 283 LEDILSTY 290
RESULT 35
US-10-335-977-7866
Sequence 7866, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 7866:
SEQUENCE CHARACTERISTICS:
LENGTH: 451 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES

ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...451
SEQUENCE DESCRIPTION: SEQ ID NO: 7866:
US-10-335-977-7866
Query Match 1.5%; Score 8; DB 12; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 63 LEDILSTY 70
Db 284 LEDILSTY 291
RESULT 36
US-09-738-626-6430
Sequence 6430, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAYAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIALI, KEIKO
APPLICANT: YOKOL, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 6430
LENGTH: 484
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-6430
Query Match 1.5%; Score 8; DB 9; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 181 AELLLEHR 188
Db 80 AELLLEHR 87
RESULT 37
US-10-369-493-23425
Sequence 23425, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493

Query Match 1.5%; Score 8; DB 15; Length 525;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
SEQUENCE ID NO 23425
LENGTH: 525
TYPE: PRT
ORGANISM: Deinococcus radiodurans
US-10-369-493-23425

QY 496 APSSPRVT 503
Db 45 APSSPRVT 52

RESULT 38
US-09-962-055-41
Sequence 41, Application US/09962055
Patent No. US20020052033A1
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/962,055
FILING DATE: 24-Sep-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/979,608
FILING DATE: 26-NOV-1997
APPLICATION NUMBER: US 60/031,930
FILING DATE: 27-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-962-055-41

Query Match 1.3%; Score 7; DB 9; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
SEQUENCE ID NO 59
LENGTH: 7

Db 1 RDVSEEL 7

RESULT 39
US-09-976-740-41
Sequence 41, Application US/09976740
Publication No. US20020194633A1
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
APPLICANT: Lees, Robert S.
APPLICANT: Law, Simon W.
APPLICANT: Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
ATHEROSCLEROSIS
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/09/976,740
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 09/616,289
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 08/979,608
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 41
LENGTH: 7
TYPE: PRT
ORGANISM: Homo sapiens
US-09-976-740-41

Query Match 1.3%; Score 7; DB 9; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+06; 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
SEQUENCE ID NO 59
LENGTH: 7

QY 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 40
US-10-671-242-41
Sequence 41, Application US/10671242
Publication No. US20040040049A1
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
APPLICANT: Lees, Robert S.
APPLICANT: Law, Simon W.
APPLICANT: Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
ATHEROSCLEROSIS
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/10/671,242
CURRENT FILING DATE: 2003-09-24
PRIOR APPLICATION NUMBER: US/09/616,289
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 09/517,849
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 08/979,608
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 41
LENGTH: 7

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-41

Query Match      1.3%; Score 7; DB 12; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      53 RDVSEEL 59
Db      1 RDVSEEL 7

RESULT 41
US-10-023-529-41
; Sequence 41, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-529-41

Query Match      1.3%; Score 7; DB 13; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      53 RDVSEEL 59
Db      1 RDVSEEL 7

RESULT 42
US-10-023-523-41
; Sequence 41, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-41

Query Match      1.3%; Score 7; DB 12; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      53 RDVSEEL 59
Db      1 RDVSEEL 7

RESULT 41
US-10-023-529-41
; Sequence 41, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-529-41

Query Match      1.3%; Score 7; DB 13; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      53 RDVSEEL 59
Db      1 RDVSEEL 7

RESULT 42
US-10-023-523-41
; Sequence 41, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-41

Query Match      1.3%; Score 7; DB 12; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      53 RDVSEEL 59
Db      1 RDVSEEL 7

RESULT 41
US-10-023-529-41
; Sequence 41, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-41

Query Match      1.3%; Score 7; DB 15; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      53 RDVSEEL 59
Db      1 RDVSEEL 7

RESULT 44
US-10-225-567A-2071
; Sequence 2071, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
; APPLICANT: Lifespan Biosciences
; APPLICANT: Brown, Joseph P.
```



```

; APPLICANT: Burner, Glenna C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2071
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-567A-2071

Query Match
Best Local Similarity 100.0%; DB 14; Length 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 BELSRQL 63
Db 6 BELSRQL 12

RESULT 45
US-10-195-730-348
; Sequence 348, Application US/10195730
; Publication No. US2003014492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P2017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,852
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 348
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-195-730-348

Query Match
Best Local Similarity 100.0%; DB 81; Length 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 427 LLEMAEE 433
Db 5 LLEMAEE 11

RESULT 46
US-09-864-761-36055
; Sequence 36055, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312

; APPLICANT: Burner, Glenna C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2071
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-567A-2071

; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/532,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36055
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO 299716.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EST HUMAN HIT: BE312629.1, EVALUATE 6.00e-15
; OTHER INFORMATION: SWISSPROT HIT: Q60429, EVALUATE 7.00e-16
; OTHER INFORMATION: SWISSPROT HIT: Q60429, EVALUATE 7.00e-16
US-09-864-761-36055

Query Match
Best Local Similarity 100.0%; DB 9; Length 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 281 KLRQENM 287
Db 23 KLRQENM 29

RESULT 47
US-09-491-614-31
; Sequence 31, Application US/09491614
; Publication No. US20030100027A1
; GENERAL INFORMATION:
; APPLICANT: Coyler, John
; APPLICANT: Lightowler, Joanne
```

```
; TITLE OF INVENTION: Methods and Compositions Using Coiled Binding Partners
; FILE REFERENCE: 4256/78978
; CURRENT APPLICATION NUMBER: US/09/491,614
; CURRENT FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: 09/259,474
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Peptide
; OTHER INFORMATION: construct for coiled coil derived from the folate
; OTHER INFORMATION: receptor and the consensus for G2i anchor
; OTHER INFORMATION: addition.
US-09-491-614-31

Query Match      1.3%; Score 7; DB 10; Length 44;
Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy      406 IKKLEKE 412
Db      23 IKKLEKE 29

RESULT 48
US-09-986-480-335
; Sequence 335, Application US/09986480
; Publication No. US20030027999A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 143 Human Secreted Proteins
; FILE REFERENCE: PS500P1
; CURRENT APPLICATION NUMBER: US/09/986,480
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: PCT/US00/12788
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/134,068
; PRIOR FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 456
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 335
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-480-335

Query Match      1.3%; Score 7; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy      12 EGAQERP 18
Db      34 EGAQERP 40

RESULT 49
US-09-892-877-355
; Sequence 355, Application US/09892877
; Publication No. US20030077809A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P1
; CURRENT APPLICATION NUMBER: US/09/892,877
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 355
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-892-877-355

Query Match      1.3%; Score 7; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy      12 EGAQERP 18
Db      34 EGAQERP 40

RESULT 50
US-09-948-783-359
; Sequence 359, Application US/09948783
; Publication No. US20030100051A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P2
; CURRENT APPLICATION NUMBER: US/09/948,783
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/231,846
; PRIOR FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: 09/892,877
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: 09/437,658
; PRIOR FILING DATE: 1999-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/09847
; PRIOR FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 60/085,093
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,094
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,105
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,180
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,927
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,906
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,924
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,922
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,921
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,923
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,925
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,928
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,920
; PRIOR FILING DATE: 1998-05-18
; NUMBER OF SEQ ID NOS: 465
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 359
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-948-783-359

Query Match      1.3%; Score 7; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy      12 EGAQERP 18
Db      34 EGAQERP 40
```

Db 34 EGAQRP 40

RESULT 51
US-10-424-599-154654
; Sequence 154654, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 154654
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(55)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_110674C.1.pep
US-10-424-599-154654

Query Match 1.3%; Score 7; DB 12; Length 55;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 PAVEAEG 29
DB 37 PAVEAEG 43

RESULT 52
US-10-424-599-246225
; Sequence 246225, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 246225
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(78)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_64373C.1.pep
US-10-424-599-246225

Query Match 1.3%; Score 7; DB 12; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 315 DLQQLV 321
DB 16 DLQQLV 22

RESULT 53
US-10-425-114-68147
; Sequence 68147, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jirongdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 68147
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 701174728_FLI.pep
US-10-425-114-68147

Query Match 1.3%; Score 7; DB 12; Length 80;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 518 GQTGPQE 524
DB 58 GQTGPQE 64

RESULT 54
US-09-796-692-1230
; Sequence 1230, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAP
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07

```
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1230
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-892-1230
```

```
Query Match      1.3%; Score 7; DB 9; Length 82;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      399 MERMTKK 405
          |||||
Db      20 MERMTKK 26
```

```
RESULT 55
US-10-040-862-1230
; Sequence 1230, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013520US
; CURRENT FILING DATE: 2001-11-06
; PRIOR FILING DATE: 2001-11-06
; PRIOR FILING DATE: 2000-03-01
; PRIOR FILING DATE: 2000-03-17
; PRIOR FILING DATE: 2000-04-27
; PRIOR FILING DATE: 2000-04-28
; PRIOR FILING DATE: 2000-05-01
; PRIOR FILING DATE: 2000-05-04
; PRIOR FILING DATE: 2000-05-22
; PRIOR FILING DATE: 2000-07-14
; PRIOR FILING DATE: 2000-08-03
; PRIOR FILING DATE: 2000-08-04
; PRIOR FILING DATE: 2000-08-07
; PRIOR FILING DATE: 2000-08-07
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1230
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-040-862-1230
```

```
Query Match      1.3%; Score 7; DB 14; Length 82;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      399 MERMTKK 405
          |||||
Db      20 MERMTKK 26
```

```
RESULT 56
US-10-057-475B-1230
; Sequence 1230, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Clapper, Jonathan David
; APPLICANT: Wang, Aijun
; APPLICANT: Ordonez, Nadia
; APPLICANT: Carter, Lauren
; APPLICANT: McNeill, Patricia Dianne
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-014402US
; CURRENT APPLICATION NUMBER: US/10/057,475B
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 10979
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1230
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-475B-1230
```

```
Query Match      1.3%; Score 7; DB 15; Length 82;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      399 MERMTKK 405
          |||||
Db      20 MERMTKK 26
```

```
RESULT 57
US-10-154-884B-1230
; Sequence 1230, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013521US
; CURRENT APPLICATION NUMBER: US/10/154,884B
```

; CURRENT FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1230
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-154-884B-1230

Query Match 1.3%; Score 7; DB 15; Length 82;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 399 MEKMTKK 405
DB 20 MEKMTKK 26

RESULT 58
US-10-424-599-197528
; Sequence 197528, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yitua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 197528
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_20394C.1.pep
US-10-424-599-197528

Query Match 1.3%; Score 7; DB 12; Length 89;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 VLARSKL 225
DB 21 VLARSKL 27

RESULT 59
US-09-796-692-1236

; Sequence 1236, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1236
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-1236

Query Match 1.3%; Score 7; DB 9; Length 91;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 399 MEKMTKK 405
DB 20 MEKMTKK 26

RESULT 60
US-09-796-692-1693
; Sequence 1693, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28

```

; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1236
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-040-862-1236

Query Match      1.3%; Score 7; DB 14; Length 91;
Best Local Similarity 100.0%; Pred.No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      399 MEMTKK 405
Db      20 MEMTKK 26

RESULT 62
US-10-040-862-1693
; Sequence 1693, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therap
; FILE REFERENCE: 014058-013520US
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1693
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-040-862-1693

Query Match      1.3%; Score 7; DB 14; Length 91;
Best Local Similarity 100.0%; Pred.No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      399 MEMTKK 405
Db      20 MEMTKK 26

RESULT 61
US-10-040-862-1236
; Sequence 1236, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013520US
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US/10/040,862
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692

```

Db

20 MEKMTXK 26

RESULT 63

US-10-057-475B-1236
; Sequence 1236, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Clapper, Jonathan David
; APPLICANT: Wang, Aijun
; APPLICANT: Ordonez, Nadia
; APPLICANT: Carter, Lauren
; APPLICANT: McNeill, Patricia Dianne

; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy

; FILE REFERENCE: 014058-014402US

; CURRENT APPLICATION NUMBER: US/10/057,475B

; CURRENT FILING DATE: 2002-01-22

; PRIOR FILING DATE: 2000-03-01

; PRIOR FILING DATE: 2000-04-27

; PRIOR FILING DATE: 2000-04-28

; PRIOR FILING DATE: 2000-05-01

; PRIOR FILING DATE: 2000-05-04

; PRIOR FILING DATE: 2000-05-22

; PRIOR FILING DATE: 2000-07-14

; PRIOR FILING DATE: 2000-08-03

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

; PRIOR FILING DATE: 2000-08-04

Query Match 1.3%; Score 7; DB 15; Length 91;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

399 MEKMTXK 405

| | | | |

20 MEKMTXK 26

RESULT 64

US-10-057-475B-1693
; Sequence 1693, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander

; APPLICANT: Algate, Paul A.

; APPLICANT: Mannion, Jane

; APPLICANT: Clapper, Jonathan David

; APPLICANT: Wang, Aijun

; APPLICANT: Ordonez, Nadia

; APPLICANT: Carter, Lauren

; APPLICANT: McNeill, Patricia Dianne

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy

; FILE REFERENCE: 014058-013521US

; CURRENT APPLICATION NUMBER: US/10/154,884B

; CURRENT FILING DATE: 2002-05-23

; PRIOR FILING DATE: 2000-03-01

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

Query Match 1.3%; Score 7; DB 15; Length 91;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

399 MEKMTXK 405

| | | | |

20 MEKMTXK 26

RESULT 65

US-10-154-884B-1236
; Sequence 1236, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander

; APPLICANT: Algate, Paul A.

; APPLICANT: Mannion, Jane

; APPLICANT: Retter, Marc W.

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy

; FILE REFERENCE: 014058-013521US

; CURRENT APPLICATION NUMBER: US/10/154,884B

; CURRENT FILING DATE: 2002-05-23

; PRIOR FILING DATE: 2000-03-01

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

; PRIOR FILING DATE: 2000-03-17

;; PRIOR APPLICATION NUMBER: US 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: US 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 11290
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 1236
;; LENGTH: 91
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-154-884B-1236

Query Match 1.3%; Score 7; DB 15; Length 91;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 399 MEKMTKK 405
DB 20 MEKMTKK 26

RESULT 66

US-10-154-884B-1693
;; Sequence 1693, Application US/10154884B
;; Publication No. US20040005561A1
;; GENERAL INFORMATION:
;; APPLICANT: Gaiger, Alexander
;; APPLICANT: Algate, Paul A.
;; APPLICANT: Mannion, Jane
;; APPLICANT: Retter, Marc W.
;; APPLICANT: Corixa Corporation
;; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;; FILE REFERENCE: 014058-013521US
;; CURRENT APPLICATION NUMBER: US/10/154,884B
;; CURRENT FILING DATE: 2002-05-23
;; PRIOR APPLICATION NUMBER: US 60/186,126
;; PRIOR FILING DATE: 2000-03-01
;; PRIOR APPLICATION NUMBER: US 60/190,479
;; PRIOR FILING DATE: 2000-03-17
;; PRIOR APPLICATION NUMBER: US 60/200,545
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: US 60/200,303
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,779
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,999
;; PRIOR FILING DATE: 2000-05-01
;; PRIOR APPLICATION NUMBER: US 60/202,084
;; PRIOR FILING DATE: 2000-05-04
;; PRIOR APPLICATION NUMBER: US 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: US 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 11290
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 1693
;; LENGTH: 91
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-154-884B-1693

Query Match 1.3%; Score 7; DB 15; Length 91;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 399 MEKMTKK 405

DB 20 MEKMTKK 26

RESULT 67

US-09-864-761-39633
;; Sequence 39633, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharron G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Aomicca-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 39633
;; LENGTH: 95
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC002465.1
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.69
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.42
US-09-864-761-39633

Query Match 1.3%; Score 7; DB 9; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEKKK 255

Db 58 EBEKEK 64

RESULT 68

US-09-804-014A-56
; Sequence 56, Application US/09804014A
; Publication No. US20030064489A1
; GENERAL INFORMATION:
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Vernet, Corine
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard
; APPLICANT: Spaderna, Steven
; APPLICANT: Majumder, Kumud
; FILE OF INVENTION: Novel Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-721 US
; CURRENT APPLICATION NUMBER: US/09/804,014A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/188,316
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 60/188,277
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 60/189,139
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: 60/189,140
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: 60/190,401
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/190,231
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 56
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-804-014A-56

Query Match 1.3%; Score 7; DB 12; Length 99;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 EBGVQRA 247

Db 2 EBGVQRA 8

RESULT 69

US-10-424-599-266203
; Sequence 266203, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; FILE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 266203
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(100)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_82402C.1.pep

US-10-424-599-266203

Query Match 1.3%; Score 7; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 ERLKXLI 297

Db 45 ERLKXLI 51

RESULT 70

US-10-424-599-153810
; Sequence 153810, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; FILE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 153810
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_109913C.1.pep
US-10-424-599-153810

Query Match 1.3%; Score 7; DB 12; Length 101;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 RAREEE 252

Db 55 RAREEE 61

RESULT 71

US-10-264-049-4293
; Sequence 4293, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; FILE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCI/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 4293
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (34)-
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (63)-
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE

; LOCATION: (106)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-4293

Query Match 1.3%; Score 7; DB 15; Length 106;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 477 QGSLTDS 483
Db 81 QGSLTDS 87

RESULT 72

US-10-195-730-345
; Sequence 345, Application US/10195730
; Publication No. US20030144492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P201721
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 345
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-195-730-345

Query Match 1.3%; Score 7; DB 14; Length 110;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 427 LLEMAEE 433
Db 81 LLEMAEE 87

RESULT 73

US-10-425-114-50441
; Sequence 50441, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 50441
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURES:
; OTHER INFORMATION: Clone ID: LIB3079-022-B2_FLI.pep
US-10-425-114-50441

Query Match 1.3%; Score 7; DB 12; Length 113;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 151 GLGKEIT 157
Db 20 GLGKEIT 26

RESULT 74

US-10-424-599-165582
; Sequence 165582, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 165582
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURES:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_120536C.1.pep
US-10-424-599-165582

Query Match 1.3%; Score 7; DB 12; Length 114;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 NRSLEKE 242
Db 45 NRSLEKE 51

RESULT 75

US-10-424-599-196307
; Sequence 196307, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 196307
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURES:
; NAME/KEY: unsure
; LOCATION: (1)..
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURES:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19282C.1.pep
US-10-424-599-196307

Query Match 1.3%; Score 7; DB 12; Length 115;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 250 EEEKKE 256
Db 97 EEEKKE 103

```

; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...117
; SEQUENCE DESCRIPTION: SEQ ID NO: 4955:
US-10-335-977-4955

Query Match      1.3%; Score 7; DB 12; Length 117;
Best Local Similarity 100.0%; Pred. No. 3e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY      406 IKKLEKE 412
DB      4 IKKLEKE 10
      |||||

RESULT 78
US-10-424-599-241987
; Sequence 241987, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 241987
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(119)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_60541C.1.pep
US-10-424-599-241987

Query Match      1.3%; Score 7; DB 12; Length 119;
Best Local Similarity 100.0%; Pred. No. 3e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY      127 EEIRQSD 133
DB      93 EEIRQSD 99
      |||||

RESULT 79
US-10-443-622-121
; Sequence 121, Application US/10443622
; Publication No. US20040024192A1
; GENERAL INFORMATION:
; APPLICANT: Carter et al.
; TITLE OF INVENTION: 19 Human Secreted Proteins
; FILE REFERENCE: P2009P1
; CURRENT APPLICATION NUMBER: US/10/443,622
; CURRENT FILING DATE: 2003-05-23
; PRIOR APPLICATION NUMBER: PCT/US98/13608
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: 60/051,480
; PRIOR FILING DATE: 1997-07-01
; PRIOR APPLICATION NUMBER: 60/051,381
; PRIOR FILING DATE: 1997-07-01
; PRIOR APPLICATION NUMBER: 60/058,663
; PRIOR FILING DATE: 1997-09-12

; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...117
; SEQUENCE DESCRIPTION: SEQ ID NO: 4955:
US-10-043-487-334

Query Match      1.3%; Score 7; DB 14; Length 115;
Best Local Similarity 100.0%; Pred. No. 2.9e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY      314 KDLQOOL 320
DB      86 KDLQOOL 92
      |||||

RESULT 77
US-10-335-977-4955
; Sequence 4955, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSER: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 4955:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Shigella flexneri
; FEATURE:
; NAME/KEY: Pierre, LEGRAIN
; TITLE OF INVENTION: Protein-protein interactions between Shigella flexneri polypeptides
; FILE REFERENCE: B4778A
; CURRENT APPLICATION NUMBER: US/10/043,487
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/261,130
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 334
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Shigella flexneri
US-10-043-487-334
```

```
; PRIOR APPLICATION NUMBER: 60/058,598
; PRIOR FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 121
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-443-622-121

Query Match      1.3%; Score 7; DB 16; Length 122;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 468 RVQDLA 474
Db 4 RVQDLA 10

RESULT 80
US-10-424-599-150351
; Sequence 150351, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kowalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 150351
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_10678C.1.pcp
US-10-424-599-150351

Query Match      1.3%; Score 7; DB 12; Length 124;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 403 TKIKKL 409
Db 17 TKIKKL 23

RESULT 81
US-08-849-303-19
; Sequence 19, Application US/08849303
; Publication No. US20030221209A1
; GENERAL INFORMATION:
; APPLICANT: Atkinson, Howard J.
; APPLICANT: McPherson, Michael J.
; APPLICANT: Urwin, Peter E.
; TITLE OF INVENTION: MODIFIED PROTEINASE INHIBITORS
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/849,303
; FILING DATE: 21-MAY-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1321-1-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 127 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-849-303-19

Query Match      1.3%; Score 7; DB 8; Length 127;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 241 EGVQRA 247
Db 27 EGVQRA 33

RESULT 82
US-10-108-260A-3437
; Sequence 3437, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560Arel full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3437
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3437

Query Match      1.3%; Score 7; DB 15; Length 130;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 488 RPEGPGA 494
Db 77 RPEGPGA 83

RESULT 83
US-10-335-977-4956
; Sequence 4956, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
```

```

NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 4957:
SEQUENCE CHARACTERISTICS:
LENGTH: 133 amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...133
SEQUENCE DESCRIPTION: SEQ ID NO: 4957:
US-10-335-977-4957

Query Match 1.3%; Score 7; DB 12; Length 133;
Best Local Similarity 100.0%; Pred.No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 IKKLEKE 412
DB 4 IKKLEKE 10

RESULT 85
US-10-335-977-4958
Sequence 4958, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 4958:
SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:

```

```

NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 4957:
SEQUENCE CHARACTERISTICS:
LENGTH: 133 amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...133
SEQUENCE DESCRIPTION: SEQ ID NO: 4957:
US-10-335-977-4957

Query Match 1.3%; Score 7; DB 12; Length 133;
Best Local Similarity 100.0%; Pred.No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 IKKLEKE 412
DB 4 IKKLEKE 10

RESULT 85
US-10-335-977-4958
Sequence 4958, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 4958:
SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:

```

```
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...137
; SEQUENCE DESCRIPTION: SEQ ID NO: 4958:
US-10-335-977-4958

Query Match      1.3%; Score 7; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 IKKLEKE 412
DB 8 IKKLEKE 14

RESULT 86
US-09-847-208-130
; Sequence 130, Application US/09847208
; Publication No. US20030082190A1
; GENERAL INFORMATION:
; APPLICANT: Saxon, Andrew
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daocheng
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
; FILE REFERENCE: U67.002A
; CURRENT APPLICATION NUMBER: US/09/847,208
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 130
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Parietaria judaica
US-09-847-208-130

Query Match      1.3%; Score 7; DB 10; Length 138;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 EKEPSKG 122
DB 58 EKEPSKG 64

RESULT 87
US-09-847-208-128
; Sequence 128, Application US/09847208
; Publication No. US20030082190A1
; GENERAL INFORMATION:
; APPLICANT: Saxon, Andrew
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daocheng
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
; FILE REFERENCE: U67.002A
; CURRENT APPLICATION NUMBER: US/09/847,208
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 128
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Parietaria judaica
US-09-847-208-128

Query Match      1.3%; Score 7; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 EKEPSKG 122
DB 58 EKEPSKG 64

US-10-425-114-69680
; Sequence 69680, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 69680
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17176H10_FLI.pep
US-10-425-114-69680

Query Match      1.3%; Score 7; DB 12; Length 142;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 SQLVQEK 208
DB 105 SQLVQEK 111

US-10-424-599-167688
; Sequence 167688, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 167688
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_122439C.1.pep
US-10-424-599-167688

Query Match      1.3%; Score 7; DB 12; Length 146;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 SNKALLE 429
DB 59 SNKALLE 65

RESULT 90
US-09-893-519A-11
; Sequence 11, Application US/09893519A
; Publication No. US20030027243A1
```


GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BUURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAO, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/1G548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 84
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Human Genbank/AA042873
; DATABASE ENTRY DATE: 2000-01-05
; RELEVANT RESIDUES: (1)..(156)
US-09-893-519A-11

Query Match 1.3%; Score 7; DB 10; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 LSRQLED 65
DB 7 LSRQLED 13

RESULT 91
US-09-934-289A-20
; Sequence 20, Application US/09934289A
; Patent No. US20020132297A1
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE
; TITLE OF INVENTION: HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: MEI098-061CF1CN1(N)
; CURRENT APPLICATION NUMBER: US/09/934,289A
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 09/342,767
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/146,950
; PRIOR FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-934-289A-20

Query Match 1.3%; Score 7; DB 9; Length 159;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 EKKKAKG 151
DB 153 EKKKAKG 159

RESULT 92
US-10-424-599-163595
; Sequence 163595, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 163595
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_118745C.1.pep
US-10-424-599-163595

Query Match 1.3%; Score 7; DB 12; Length 166;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 DRDHRRP 143
DB 71 DRDHRRP 77

RESULT 93
US-10-424-599-241739
; Sequence 241739, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 241739
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(167)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_60317C.1.pep
US-10-424-599-241739

Query Match 1.3%; Score 7; DB 12; Length 167;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 PQEKKKA 149

```
Db      85 PQKCKKA 91
|||||
RESULT 94
US-10-424-599-262163
; Sequence 262163, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 262163
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_78755C.1.pep
US-10-424-599-262163

Query Match      1.3%; Score 7; DB 12; Length 171;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      90 EPEDAEX 96
|||||
Db      130 EPEDAEX 136

RESULT 95
US-09-847-208-129
; Sequence 129, Application US/09847208
; Publication No. US20030082190A1
; GENERAL INFORMATION:
; APPLICANT: Saxon, Andrew
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daocheng
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
; TITLE OF INVENTION: IGE-MEDIATED ALLERGIC DISEASES
; FILE REFERENCE: UC67.002A
; CURRENT APPLICATION NUMBER: US/09/847,208
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 129
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Parietaria judaica
US-09-847-208-129

Query Match      1.3%; Score 7; DB 10; Length 176;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      116 EKEPSKG 122
|||||
Db      59 EKEPSKG 65

RESULT 96
US-10-276-774-2467
; Sequence 2467, Application US/10276774
; Publication No. US2004005345A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y. Tom et al
```

```
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2467
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(186)
; OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-2467

Query Match      1.3%; Score 7; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      477 QGSLTDS 483
|||||
Db      47 QGSLTDS 53

RESULT 97
US-09-934-289A-18
; Sequence 18, Application US/09934289A
; Patent No. US20020132297A1
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE
; TITLE OF INVENTION: HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: PBIO98-061CPCINI(M)
; CURRENT APPLICATION NUMBER: US/09/934,289A
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 09/342,767
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/146,950
; PRIOR FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(38)
US-09-934-289A-18

Query Match      1.3%; Score 7; DB 9; Length 197;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      145 EKKVAKG 151
|||||
Db      191 EKKVAKG 197

RESULT 98
US-10-425-114-45419
; Sequence 45419, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
```

```
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 45419
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3088-030-A1_FLI.pep
US-10-425-114-45419

Query Match      1.3%; Score 7; DB 12; Length 206;
Best Local Similarity 100.0%; Pred.No. 4.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 KOEMERK 402
DB 115 KOEMERK 121

RESULT 99
US-10-104-047-2607
; Sequence 2607, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2607
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2607

Query Match      1.3%; Score 7; DB 15; Length 207;
Best Local Similarity 100.0%; Pred.No. 4.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 KAKGLGK 154
DB 126 KAKGLGK 132

RESULT 100
US-10-424-599-244841
; Sequence 244841, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 244841
; LENGTH: 208
; TYPE: PRT
```

```
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_63120C.1.pep
US-10-424-599-244841

Query Match      1.3%; Score 7; DB 12; Length 208;
Best Local Similarity 100.0%; Pred.No. 5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 RAREEES 252
DB 44 RAREEES 50

Search completed: June 7, 2004, 14:48:11
Job time : 46 secs
```